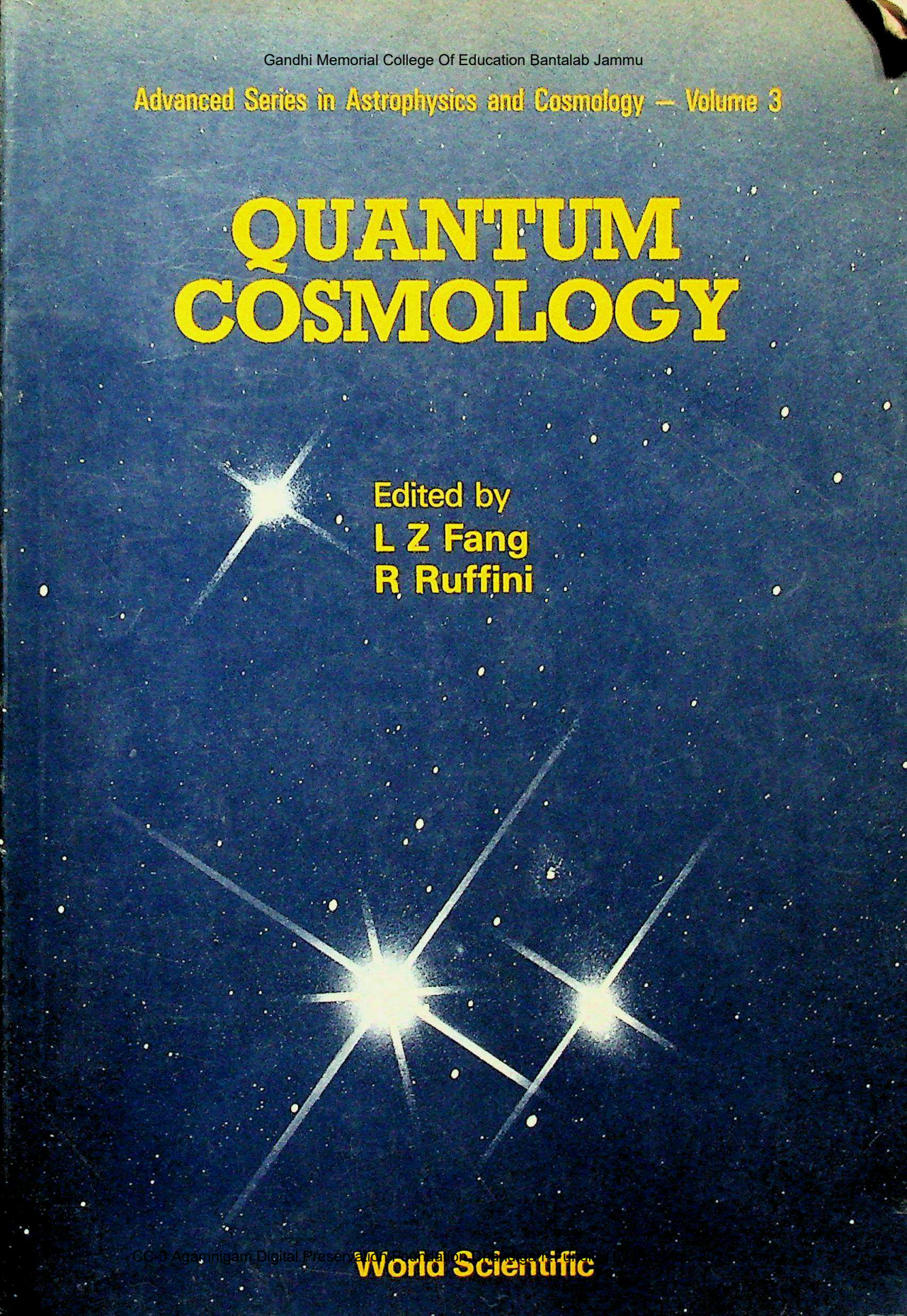
Gandhi Memorial College Of Education Bantalab Jammu

Advanced Series in Astrophysics and Cosmology — Volume 3

QUANTUM COSMOLOGY

Edited by
L Z Fang
R Ruffini



CC-0 Agamnigam Digital Preservation Foundation,New Delhi. http://www.agamnigam.com

World Scientific

Gandhi Memorial College Of Education Bantalab Jammu

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

QUANTUM COSMOLOGY

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ADVANCED SERIES IN ASTROPHYSICS AND COSMOLOGY

edited by: L Z Fang (*Beijing Astronomical Observatory*) and R Ruffini (*Roma*)

Volume 1 Cosmology of the Early Universe

Volume 2 Galaxies, Quasars and Cosmology

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Advanced Series in Astrophysics and Cosmology — Volume 3

QUANTUM COSMOLOGY

J.M. College of Education
Raipur, Bantalab
Jammu.
Acc No....5263....
Dated....19/3/2024

Edited by
Li Zhi Fang
Beijing Astronomical Observatory
Chinese Academy of Sciences
Remo Ruffini
Istituto di Fisica G Marconi
Universita degli Studi — Roma

World Scientific
Singapore • New Jersey • Hong Kong

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Published by

World Scientific Publishing Co. Pte. Ltd.
P.O. Box 128, Farrer Road, Singapore 9128

U. S. A. office: World Scientific Publishing Co., Inc.
687 Hartwell Street, Teaneck NJ 07666, USA

Library of Congress Cataloging-in-Publication data is available.

QUANTUM COSMOLOGY

Copyright © 1987 by World Scientific Publishing Co Pte Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

ISBN 9971-50-293-3
9971-50-312-3 pbk

Printed in Singapore by Kyodo-Shing Loong Printing Industries Pte Ltd.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PREFACE

The birth of the universe is a challenging problem for physicists. In the light of the recent development of cosmology in the GUT epoch, the problem of how the universe emerged from Big Bang singularity can no longer be neglected. An important approach to this problem is the wave function proposal in which the universe would be completely self-contained without any singularities and without any edges.

The aim of this volume is to provide a detailed survey of quantum cosmology under the scheme of the wave function proposal. The topics cover inflationary universe; quantum gravity; wave function of the universe; solution of toy models; and the consequences of quantum cosmology.

We would like to express our special thanks to Dr. Wu Zhong Chao for his valuable suggestions.

L. Z. Fang

R. Ruffini

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

CONTENTS

Preface v

1 **INTRODUCTION**

An Overview of Quantum Cosmology 1
L Z Fang & Z C Wu

2 **QUANTUM GRAVITY**

Superspace and the Nature of Quantum Geometrodynamics 27
J A Wheeler

Quantum Theory of Gravity. I. The Canonical Theory 93
B S DeWitt

3 **THE EARLY UNIVERSE**

A New Type of Isotropic Cosmological Models Without Singularity 130
A A Starobinsky

Cosmological Baryon-number Domain Structure and the First Order Phase Transition of a Vacuum 134
K Sato

Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems 139
A H Guth

A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems 149
A D Linde

Supercooled Phase Transitions in the Very Early Universe 154
S W Hawking & I G Moss

Cosmology for Grand Unified Theories With Radiatively Induced Symmetry Breaking 158
A Albrecht & P J Steinhardt

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

4 QUANTUM COSMOLOGY

The Boundary Conditions of the Universe 162
S W Hawking

Wave Function of the Universe 174
J B Hartle & S W Hawking

Quantum Cosmology 190
S W Hawking

5 TOY MODELS

The Quantum State of the Universe 236
S W Hawking

Higher Derivatives in Quantum Cosmology (I). The isotropic case 256
S W Hawking & J C Luttrell

Numerical Calculations of Minisuperspace Cosmological Models 267
S W Hawking & Z C Wu

6 CONSEQUENCES

The Isotropy of the Universe 273
S W Hawking & J C Luttrell

Origin of Structure in the Universe 277
J J Halliwell & S W Hawking

Quantum Cosmology with a Positive-Definite Action 292
G T Horowitz

Dimension of the Universe 301
Z C Wu

Arrow of Time in Cosmology 308
S W Hawking

Operator Ordering and the Flatness of the Universe 315
S W Hawking & D N Page

Formation of Black Holes in Quantum Cosmology 327
L Z Fang & M Li

Acknowledgements 330

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

QUANTUM COSMOLOGY

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

International Journal of Modern Physics A, Vol. 1, No. 4 (1986) 887–912
© World Scientific Publishing Company

AN OVERVIEW OF QUANTUM COSMOLOGY

L. Z. FANG*
Institute for Advanced Study, Princeton, NJ 08540, USA

Z. C. WU*
Department of Physics, Syracuse University, Syracuse, NY 13210, USA

Received 2 June 1986

Hawking's theory of quantum cosmology is the most important stage in understanding our universe since the big bang model. In principle, one can predict everything in the universe solely from physical laws. All main results within this framework have been reviewed in this paper.

1. Introduction

The standard Big Bang cosmological model is a successful theory to explain the most important astronomical observations, namely, Hubble's redshsift law, the cosmic helium abundance and the cosmic microwave background. It is estimated that the big bang should have occurred between 10 and 20 billion years ago.

In this model the whole evolution of the universe crucially depends on the initial conditions chosen near the big bang singularity. Unless these initial conditions are chosen in a very special way, it is impossible to overcome many puzzles, as for instance, the horizon problem, the flatness problem and the monopole problem. In order to solve these problems it has been suggested that a first order phase transition would have occurred in the GUT era, about 10^{-35} s after the big bang. Before this phase transition, the temperature was very high, the universe was in a false vacuum state, and the energy density behaved as an effective positive cosmological constant which caused the universe to expand exponentially. Within any observer's horizon, most fluctuations in the fields present before the inflation would have been smeared away by the expansion. Thus the quantum fluctuations of fields in the inflation stage became the only origin for the later development of the perturbations, leading to the formation of galaxies and all other structures of the universe. It turns out that a rather wide range of initial conditions would have led to the same universe we live in.

The new developments in quantum cosmology are the most important stage in understanding the birth of the universe. The old motivation of quantum cosmology in the sixties was just to smear out the big bang singularity predicted in classical general

* Permanent address: Center for Astrophysics, University of Science and Technology of China, Hefei, Anhui, People's Republic of China.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

relativity, exactly in the same way as quantum mechanics prevents the orbiting electrons in an atom from falling into the nucleus. However, the most important task in cosmology is to discover the manner by which the universe was born. Cosmology, as a branch of theoretical physics must not be only self-consistent, but also self-contained. Since there is nothing outside the universe, one can rephrase this by stating that the boundary condition of the universe is that there is no boundary. Such a self-contained theory has been found in a recent presentation of the ground state proposal by Hawking and Hartle: the quantum state of the universe must be defined by a path integral summed over all compact 4-metrics.

The wave function of the universe has arbitrary closed space-like 3-geometries and matter fields on them as the configuration space. As in classical general relativity, the quantum state does not depend on the coordinate description and this fact compels the wave function to obey the Wheeler-DeWitt equation, i.e., the zero-energy Schrödinger equation. Just as in the case of a particle moving in a potential well, which classically can be found only where the quantum mechanical wave function has an oscillating behavior, the oscillatory part of the wave function of the universe is interpreted as an ensemble of classical Lorentzian evolutions. It is believed that any realistic classical evolution must be included in the wave packet. The concept of time and the Lorentzian evolution become meaningful as soon as the wave function starts to oscillate.

It is very difficult to do calculations in quantum cosmology except for some simple models. The Hawking massive scalar model seems to be the most promising one until now. It has been shown that in the case of a homogeneous isotropic closed universe with a massive scalar field ϕ the wave function corresponds in the classical limit to a family of classical solutions which have a long period of exponential or inflationary expansion around the Planck time and then go over to the matter-dominated expansion of the big bang model. Perturbation calculations around this mini-superspace background carried out by Halliwell and Hawking show that the spectrum of density perturbation is scale free, which is needed to account for the origin of galaxies and all other structures in the universe. Furthermore, Hawking argues that the observed time asymmetry or the arrow of time defined by the direction in which entropy increases is related to the cosmological arrow of time defined by the direction of the expanding universe. This arises because the universe would have been smooth and homogeneous when it was small, but became disordered and inhomogeneous when it became large.

Some alternative models have also been studied. In the model with higher derivative quantum corrections, the wave function takes on a similar form as that of the Hawking model. To study the anisotropy of the universe, it is found that most of the anisotropy of the Bianchi-IX universe, the only closed, homogeneous but anisotropic cosmological model, is damped away by the Planck inflation. To deal with the flatness problem, one needs to define the probability of finding a 3-geometry in quantum cosmology, which is equivalent to eliminating the operator-ordering ambiguity in the Wheeler-DeWitt equation or the measure ambiguity in the path-integral approach. Hawking and Page have proposed the following ansatz: the Wheeler-DeWitt equation must be coordinate-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

invariant with respect to the supermetric in configuration space. As an application, they show that the probability distribution is entirely concentrated at the critical value of the matter density. In general, the most probable evolution of the universe arises from some gravitational instanton by quantum tunneling. One of the goals of the study of these instanton solutions is to show that they lead in a natural way to a 4-dimensional space-time. Within the framework of 11-dimensional supergravity as a true Kaluza-Klein theory, by using some mini-superspace ansatz, it is proved that only $S^4 \times S^7$ instantons are possible. It seems that the main reason why the apparent space-time is 4-dimensional is simply that there does not exist any gravitational instanton without a 4-dimensional sector in the theory.

These results indicate that the new developments of quantum cosmology have superseded the inflationary universe.

One can also discuss models with different topologies, include fermion fields and even study supersymmetric models and the primordial black hole problem: all these aspects are presently under investigation. In principle, one can predict everything in the universe solely from physical laws. Thus, the long-standing "first cause" problem intrinsic in cosmology has been finally dispelled.

This article is structured as follows. Section 2 is devoted to a short review of the inflationary universe. We review the canonical formulation of general relativity and quantum gravity in Secs. 3 and 4, respectively. The Hartle-Hawking ground state proposal is presented in Sec. 5. The de Sitter model, the Hawking massive scalar model, the higher derivative model are discussed in Secs. 6, 7 and 8, respectively. The isotropy, origin of the structure, time arrow, flatness and dimensions of the universe are investigated in the remaining five sections.

2. The Inflationary Universe

The hot big bang model is a good theory to describe our universe.[1] It explains the microwave-background radiation, the cosmic helium abundance and the Hubble red-shift law. In this model the space-time metric takes a Friedman-Robertson-Walker form

$$ds^2 = -dt^2 + R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\, d\Omega^2\right], \tag{2.1}$$

where $k = 1, 0, -1$ corresponds to closed, flat or hyperbolic open universe, respectively, and $R(t)$ is the scale factor of the comoving 3-surface.

Since the Einstein equation is hyperbolic, one must specify initial conditions to obtain solutions of it. In general, these conditions are assumptions put in by hand. It turns out that unless these initial conditions are chosen in a very special way, it is impossible to explain the following problems.

(i) Horizon problem:[2] why is the universe so homogeneous and isotropic even for regions which were never in causal contact with each other?

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

(ii) Flatness problem:[3] why is the matter density very near the critical value which separates the recollapse of the universe from its hyperbolic expansion to an incredible accuracy at very early times?

(iii) Monopole problem:[4] In the GUT era, the very high temperature of the order of 10^{15} GeV makes strong, weak and electromagnetic interactions indistinguishable. Following the expansion of the universe, the temperature dropped down and spontaneous symmetry breaking occurred. Monopoles should be produced as a result of knots of different symmetry breaking regions with different directions, since these regions were out of causal contact before the onset of spontaneous symmetry breakdown. However, the predicted number of monopoles formed in this way is unacceptably large, in view of the lack of evidence for the existence of monopoles.

The evolution of the universe is governed by the Einstein field equation and it has been generally assumed that the entropy must be conserved during this evolution. Therefore, at the radiation-dominated stage, the scale factor of the universe grew like $t^{1/2}$ and the above-mentioned puzzles are unavoidable. Many authors realized that in the GUT era a first order phase transition would have occurred.[5] The universe started off approximately in thermal equilibrium at the Planck temperature of 10^{19} GeV and cooled as it expanded. When the temperature was much higher than 10^{15} GeV the effective potential $V(\phi)$ of the Higgs scalar field took its minimum value only at the symmetric phase $\phi = 0$. Since the effective potential includes the quantum and thermal corrections, when the temperature fell below 10^{15} GeV, new minima of the potential at the symmetry breaking phase formed and the symmetric phase became metastable. In this case the potential of the symmetric phase appeared as a positive cosmological constant which led the universe to expand exponentially in the de Sitter manner.

The quantum tunneling from the original symmetric phase to the symmetry breaking phase was supposed to have occurred through the formation of bubbles surrounded by the symmetric phase.[6] Once formed, the bubble walls expanded hyperbolically and the volume energy of the symmetric false vacuum was transferred into the kinetic energy of the wall. Following the exponential expansion of the background space-time, the temperature fell to a very low value. The collision of bubbles was supposed to have reheated the universe to roughly the original GUT temperature. Because at the reheating stage the entropy was not conserved, if the supercooled expansion stage was long enough, all three of the puzzles can be resolved.

However, since the formation and the expansion of bubbles occurred in the expanding de Sitter background, the probability for the collisions of more than a few bubbles is extremely small. This would have produced a very inhomogeneous universe which is not the one we live in.[7] It was then suggested that the potential of the Higgs scalar field must take the Coleman-Weinberg form[8]

$$V(\phi) = \frac{\alpha^2}{2}\left[\phi^4 \ln^2(\phi\sigma^{-1}) - \frac{\phi^4}{2} + \frac{\sigma^4}{2}\right], \tag{2.2}$$

where α is the gauge coupling constant. Near the symmetric phase $\phi = 0$ the potential

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

is very flat. It took a very long time for ϕ to roll down the potential hill. After the rolling down, the scalar field oscillated around the minimum $\phi = \sigma$. The oscillation can be thought of as a large number of Higgs particles in a coherent state which decayed into the radiation of the big bang model. During the supercooled expansion the metric can be approximated by a de Sitter space. By the No Hair theorem of de Sitter space,[9] any fluctuation except for black holes before this stage would have been smeared away by the expansion and the quantum fluctuations of de Sitter space is the only source for later perturbations, such as those leading to the formation of galaxies. The evolution of the fluctuations can be described in terms of a spatial variation in the position of the scalar field on the potential hill as it rolls down to the minimum.[10] One can use a harmonic expansion on the four sphere S^4, the Euclidean version of de Sitter space, to work out the magnitude of the quantum fluctuation. The result should be compatible with observations. If the fluctuation was too large, then gravitational collapse would have occurred.[11] From a series of calculations, it is found that in such a new inflationary universe, the spectrum of the later perturbations is scale-free, as expected by Zel'dovich and Harrison,[12] but unfortunately—with too large a magnitude. However, in the supersymmetric GUTs, the magnitude of the perturbation would be smaller because the fermion loops contribute with the opposite sign to the vector loops to reduce the effective self-coupling of the Higgs field. In this case the predicted fluctuations might be in agreement with the observations.

3. Canonical Formulation of General Relativity

In the inflationary universe we have always implicitly assumed that before the GUT phase transition the universe was in the symmetric phase. To justify this assumption one should study the universe at even earlier times. At the Planck era, $\sim 10^{-44}$ sec., the quantum gravity effects would have played the most important role, so we must first discuss quantum gravity.[13]

Any space-time metric can be written locally in the form

$$ds^2 = -(N^2 - N_i N^i)\,dt^2 + 2N_i\,dx^2 + h_{ij}\,dx^i\,dx^j, \tag{3.1}$$

where N and N_i are, respectively, the lapse function and the shift vector. The 3-surfaces of constant t embedded in space-time are described by the intrinsic metric h_{ij} and the external curvature is

$$K_{ij} = \frac{1}{2N}\left[-\frac{\partial h_{ij}}{\partial t} + 2N_{(i|j)}\right] \tag{3.2}$$

The gravitational action is written as

$$I_g = \int L_g\,d^3x\,dt = \int \frac{N}{16\pi}(G^{ijkl}K_{ij}K_{kl} + h^{1/2}R)\,d^3x\,dt, \tag{3.3}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where

$$G^{ijkl} = \frac{1}{2}h^{1/2}(h^{ik}h^{jl} + h^{il}h^{jk} - 2h^{ij}h^{kl}), \tag{3.4}$$

is called the DeWitt metric (or supermetric) with signature $(-+++++)$ for each point on the 3-surfaces.

The Hamiltonian is

$$H = \int (\pi^{ij}h_{ij} - L_g)\, d^3x = \int (NH_0 + N_iH^i)\, d^3x, \tag{3.5}$$

where the canonical conjugate momenta are

$$\pi^{ij} = \frac{\partial L_g}{\partial \dot{h}_{ij}} = -\frac{h^{1/2}}{16\pi}(K^{ij} - h^{ij}K), \tag{3.6}$$

and

$$H_0 = 16\pi G_{ijkl}\pi^{ij}\pi^{kl} - \frac{1}{16}h^{1/2}\,{}^3R. \tag{3.7}$$

$$H^i = -2\pi^{ij}_{|j}. \tag{3.8}$$

Here G_{ijkl} satisfies

$$G_{ijkl}G^{ijk'l'} = \delta^{k'}_{(k}\delta^{l'}_{l)}. \tag{3.9}$$

From the Hamiltonian one can derive the constraints and time development equations

$$H^i = 0, \tag{3.10}$$

$$H_0 = 0, \tag{3.11}$$

$$\dot{h}_{ij} = \frac{\partial H}{\partial \pi^{ij}}, \tag{3.12}$$

$$\dot{\pi}^{ij} = -\frac{\partial H}{\partial h_{ij}}. \tag{3.13}$$

The matter field can also be included into the theory by adding the corresponding terms in the above formulas.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

4. Quantum Gravity

In quantum gravity, all space-like 3-geometries can be taken as the configuration space. In this paper we shall restrict ourselves to the case of a closed 3-geometry being the configuration of the wave function. To quantize the gravitational field the classical variables are replaced by their corresponding operators in the representation of the 3-geometry. We then have

$$\pi^{ij}(x) \to -i\frac{\delta}{\delta h_{ij}(x)}, \quad h_{ij}(x) \to h_{ij}(x). \tag{4.1}$$

The quantum state does not depend on the time label t of the 3-geometry, so the wave function Φ obeys the zero-energy Schrödinger equation[14]

$$\hat{H}\Phi = 0. \tag{4.2}$$

Since one can arbitrary choose N and N_i, this equation can be decomposed into the Wheeler-DeWitt equation

$$\hat{H}_0\Phi = 0 \tag{4.3}$$

and the momentum constraint

$$\hat{H}^i\Phi = 0. \tag{4.4}$$

There is some ambiguity in the choice of operator ordering in the substitution (4.1), since the Wheeler-DeWitt equation is a second order equation. This ambiguity will not affect most of our discussion below except for the definition of the probability of finding a 3-geometry in the configuration space. Hawking and Page[15] had advanced an ansatz to eliminate this ambiguity and apply it to the study of the flatness of the universe. We shall take up this problem later.

One can also formulate quantum gravity by the path-integral approach.[16] The configuration of the wave function is a space-like 3-geometry and matter fields on it. The transition amplitude from the configuration h'_{ij}, ϕ' to h''_{ij}, ϕ'' can be expressed as

$$\langle h''_{ij}, \phi'' | h'_{ij}, \phi' \rangle = \int \delta[g_{\mu\nu}]\delta[\phi]\exp(iI[g_{\mu\nu}, \phi]), \tag{4.5}$$

where I is the action. The path integral is summed over all 4-manifolds and matter fields on them sandwiched by the two boundaries h'_{ij}, ϕ' and h''_{ij}, ϕ''. The time labels of the two 3-geometries h'_{ij}, h''_{ij} does not appear explicitly, since the time lapse is somehow to be defined by these 4-manifolds. In order for the path integral to be convergent, one

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

can evaluate the path integral in the Euclidean regime, i.e., by a summation of the Euclidean 4-metrics and matter fields on them.

The amplitude (4.5) can be thought of as the wave function of configuration h''_{ij}, ϕ'' of a state which evolves from the configuration h'_{ij}, ϕ'. Of course it must satisfy the Wheeler-DeWitt equation and the momentum constraints.

5. The Ground State Proposal

Not all solutions of the Wheeler-DeWitt equation can be used to describe the quantum state of the universe. It is believed that it must be described by the ground state of Hartle and Hawking.[17] That is, the wave function of the universe is defined by a path integral over all compact Euclidean metrics and matter fields on them which have the configuration h_{ij} and ϕ_0 as the only boundary

$$\Phi(h_{ij}, \phi_0) = \int_c \delta[g_{\mu\nu}]\delta[\phi]\exp(-\bar{I}[g_{\mu\nu}, \phi]), \tag{5.1}$$

where $\bar{I} = -iI[\phi]$ is called the Euclidean action. By comparing expressions (5.1) and (4.5), one can say that the universe is quantum mechanically created from nothing. Recall that in Minkowski space-time, any quantum state evolved from a configuration at the minus infinity of the imaginary time must be a ground state. State (5.1) can be called as the ground state in a similar manner.

The Euclidean gravitational action is

$$\bar{I} = -\frac{1}{16}\left[\int_M (R - 2\Lambda)g^{1/2}\, d^4x + \int_{\partial M} 2Kh^{1/2}\, d^3x\right]; \tag{5.2}$$

one can rewrite it as the Euclidean version of (3.3). Since the DeWitt metric is not positive definite,[13] one can make the action arbitrarily negative by choosing a conformal factor which is rapidly varying. To make the path integral convergent, one can use the Gibbons-Hawking-Perry prescription.[18] This is to divide the space of all metrics $g_{\mu\nu}$ into equivalence classes under conformal transformations. In each equivalence class one can rotate the contours of integration over the conformal factor at each point so that it is parallel to the imaginary axis, then the action becomes positive definite and the Euclidean functional integral is convergent. Recently, this prescription for linear gravity has been proved[19] to be equivalent to the following procedure: (i) expressing the theory in terms of its physical degrees of freedom; (ii) then formulating the quantum sum over histories in terms of these degrees of freedom, and (iii) finally adding back in integrations over redundant vairables to manifestly express the invariance of the theory.

The original Euclidean path integral is, formally at least, manifestly positive, whereas the revised prescription allows the wave function to oscillate. The oscillations turn out to be important for a classical Lorentzian space-time interpretation of the wave

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

function, so one may say the observed indefinite metric signature of our universe is a consequence of the indefinite Einstein-Hilbert action.[20]

The oscillating component of the wave function can be interpreted by the WKB ansatz

$$\Phi = C \cos S, \tag{5.3}$$

where S is a rapidly varying phase and C is a slowly varying amplitude. In the WKB approximation, the Wheeler-DeWitt equation for the wave packet can be reduced to the Hamilton-Jacobi equation with S identified as the classical action. Thus the wave packet can be interpreted as an ensemble of classical Lorentzian trajectories which satisfy the equation[17]

$$\nabla_i S = p_i, \tag{5.4}$$

where p_i are the canonical momenta.

The ground-state proposal gives very strong restriction to classical evolutions. Indeed, the second order Einstein equation has been truncated into the first order equation (5.4) by this proposal. In a model with n variables, the general evolution has $2n - 2l$ degrees of freedom, where l denotes the number of the Einstein and momentum constraints, since these constraints represent the coordinate (or gauge) invariance of the system, the gauge-independent degrees of freedom must be reduced to $2n - 2l$. This number is consistent with the considerations in ordinary canonical mechanics; the action takes both ends of the trajectories as the configuration. While for the phase S of the packet, the south pole of the 4-metrics can be seen as an end point. Thus, the number of degrees of freedom of the classical evolution is reduced to $n - l$, or, by half. Therefore, in the oscillating region, in general, there is one and only one trajectory through each point.[21]

The region of configuration space where the wave function does not take a wave packet form is called the Euclidean region. In this region the time concept breaks down and the classical evolution becomes meaningless.

No general proposal for the case of the noncompact 3-geometry as the configuration in quantum cosmology is available. Hartle[22] has generalized the ground state proposal to the case of an asymptotically flat 3-geometry with matter on it as the configuration. Then the class of the path integral C contains all asymptotically flat 4-metrics with matter on them. It turns out that for a flat space-time background this definition of the ground state is consistent with the minimum energy state in quantum field theory.

6. The de Sitter Model

The Wheeler-DeWitt equation is a second order hyperbolic differential equation on an infinite dimensional manifold-superspace. In order to solve it we should truncate the equation to a finite dimensional submanifold called mini-superspace. In the semi-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

classical approximation, the wave function (5.1) can be expressed in the form

$$\Phi(h_{ij}, \phi) = N_0 \sum_i A_i \exp(-B_i), \tag{6.1}$$

where N_0 is a normalization constant, B_i are the Euclidean actions of classical solutions which are compact and have the given 3-metric h_{ij} and matter field ϕ as a boundary, and A_i represent the quantum fluctuations about these solutions.

The simplest model is the de Sitter model. Its Euclidean metric is

$$ds^2 = \frac{2}{3\pi}[d\tau^2 + a^2(\tau)\, d\Omega_3^2], \tag{6.2}$$

where $d\Omega_3^2$ is the metric of a 3-sphere with unit radius. The action is

$$\bar{I} = \frac{1}{2}\int d\tau \frac{1}{a}\left[-\left(\frac{ada}{d\tau}\right)^2 - a^2 + \frac{\Lambda}{3}a^4\right] \tag{6.3}$$

and the Wheeler-DeWitt equation is

$$\frac{1}{2}\left[\frac{1}{a^p}\frac{\partial}{\partial a}\left[a^p \frac{\partial}{\partial a}\right] - a^2 + \frac{\Lambda}{3}a^4\right]\Phi[a] = 0, \tag{6.4}$$

where the index p represents some but not all of the ambiguity in the factor ordering in the Wheeler-DeWitt equation, or equivalently, the measure ambiguity in the path integral.

The semi-classical approximation to the path integral will be given by the solution of the classical field equations with a Λ term. If $a < H^{-1} \equiv \sqrt{\frac{3}{\Lambda}}$, the solution is the southern part of a 4-sphere of radius H^{-1} which has the 3-sphere of radius a as a boundary, and we have

$$\Phi(a) \simeq N_0 \exp[-\bar{I}(a)] = N_0 \exp\left[\frac{1}{3H^2}[1 - (1 - H^2a^2)^{3/2}]\right] \quad (a < H^{-1}). \tag{6.5}$$

One can use the joint condition of the standard WKB method to find the wave function for $a > H^{-1}$

$$\Phi(a) \simeq N_0' \exp\left(\frac{1}{3H^2}\right)\cos\left[(H^2a^2 - 1)^{3/2}/3H^2 - \frac{\pi}{4}\right] \quad (a > H^{-1}). \tag{6.6}$$

Our simple derivation for the case $a > H^{-1}$ lacks some physical intuitions. Hartle and Hawking use the K-representation to find the behaviour of the wave packet described by (6.6). The quantum amplitude can be thought of as the contribution from a couple

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

of complex 4-spheres with radius H^{-1}. The oscillating phase classically corresponds to the collapsing and expanding phases of the de Sitter space. The de Sitter model is very useful for the discussion of more complicated models.

7. The Hawking Massive Scalar Model

The second model we shall discuss is Hawking massive scalar field model.[23] The Wheeler-DeWitt equation is

$$\frac{1}{2}\left[\frac{1}{a^p}\frac{\partial}{\partial a}\left[a^p\frac{\partial}{\partial a}\right] - a^2 - \frac{\partial^2}{a^2\partial\phi^2} + a^4m^2\phi^2\right]\Phi[a,\phi] = 0. \tag{7.1}$$

Its configuration space can be divided into two regions. (a) The Euclidean region: $V \equiv a^4m^4\phi^2 - a^2 < 0$. The significance of this inequality is that the radius of the 3-sphere a is smaller than the radius of a 4-sphere with an effective cosmological constant $3m^2\phi^2$. If $m\phi \gg 1$ the Euclidean solution of the field equation must be a distorted 4-sphere and the scalar field ϕ does not change significantly on it. Then the wave function must take the form (6.5) with H^2 replaced by m^2a^2. (b) The Lorentzian region: $V > 0$. If $m\phi \gg 1$ the wave function is also similar to (6.6) but with $3m^2\phi^2$ as a cosmological constant. For the case of $m\phi < 1$, one needs to numerically integrate the Wheeler-DeWitt equation.

Equation (7.1) can be rewritten in a more convenient form by introducing new coordinates

$$x = a\sinh\phi, \quad y = \cosh\phi. \tag{7.2}$$

It then becomes ($p = 1$)

$$\frac{1}{2}\left[\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial x^2} + V(x,y)\right]\Phi = 0, \tag{7.3}$$

where $V(x,y) = x^2 - y^2 + m^2(y^2 - x^2)^2 \operatorname{arctanh}^2(x/y)$. The configuration space is the future of the null cone of the coordinate origin. The boundary condition at the null cone for equation (7.3) can be set as

$$\Phi = 1. \tag{7.4}$$

This is because near the null cone the size of the 3-sphere becomes very small and the gravitational action tends to zero; by the path-integral definition of the ground state, the wave function must be constant there. Since the boundary is the null cone, condition (7.4) is sufficient to evolve (7.3), and we do not need the first order derivative condition there. In general, it is very difficult to evaluate the wave function by the path-integral definition. In order to solve it, one should use the Wheeler-DeWitt equa-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

tion; in this way the ground state proposal reduces to the boundary condition for the Wheeler-DeWitt equation.

The numerical calculations show that Φ increases exponentially in the region with $V < 0$ and oscillates when $V > 0$ as predicted by the semi-classical approximation. These oscillations spread to the central channel in which $V < 0$. An analytic method to find approximate wave function has been developed recently by Page.[20]

One can consider the wave packet as an ensemble of classical trajectories. They start off with zero initial velocity from the space-like segment of the zero-potential surface $V = 0$ in configuration space and experience a period in which a is exponentially increasing and ϕ is steadily decreasing. Then ϕ passes through zero and starts oscillating while a increases as in a matter-dominated universe, reaches a maximum value and decreases to another bounce. The exponential growing component of the wave function can be interpreted as corresponding to a positive definite 4-metric in the classical limit.

Another mini-superspace model with an inflationary period has been constructed by Moss and Wright.[24]

8. The Higher Derivative Model

The quantum theory of gravity with the Einstein-Hilbert action is not renormalizable. One way to make it renormalizable is to include curvature squared quantum corrections to the action. Because of the Gauss-Bonnet theorem, the general action with squared curvature terms is

$$I = -\frac{1}{16\pi}\int d^4x\sqrt{g}[R - \alpha C^{\mu\nu\rho\sigma}C_{\mu\nu\rho\sigma} + \beta R^2] + \text{surface term} \quad (\alpha, \beta \text{ const}), \quad (8.1)$$

where $C_{\mu\nu\rho\sigma}$ is the Weyl tensor. A homogeneous isotropic universe with this action has the Hamiltonian[23,25]

$$H = \frac{1}{4}P_x^2 - \frac{1}{4}P_y^2 + V(x, y), \quad (8.2)$$

where $x = 2\beta aR$, $y = 2a(1 + \beta R)$ and $V = -\frac{1}{4}(y^2 - x^2) + (x^2/72\pi\beta)(x - y)^2$. The classical equations of the model can be derived from the Hamiltonian in the normal way and the zero-energy Schrödinger equation can be obtained by the usual substitutions $P_x \to -i(\partial/\partial x)$ and $P_y \to -i(\partial/\partial y)$,

$$\left[\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial x^2} + 4V(x, y)\right]\Phi = 0. \quad (8.3)$$

Equation (8.3) should be solved on the half-space $y > x$, which corresponds to $a > 0$. The action of Euclidean solutions for points near the null line $y = x(a = 0)$ is very small. Thus the boundary condition is $\Phi = 1$. The line $y = x$ is not a Cauchy surface. It should be supplemented by a knowledge of Φ on the other null line through the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

origin, i.e., $y = -x$. From the semi-classical estimation, one finds

$$\Phi \simeq \exp\left[-\frac{a^4}{36\pi\beta}\right]. \tag{8.4}$$

One can numerically integrate (8.3) from these boundary conditions. As predicted from the semi-classical approximation, the wave function starts to oscillate near the right-hand branch of the curve $V = 0$. This behaviour is very similar to that of the massive scalar model. Again the oscillating part of the wave function can be interpreted via the WKB approximation as corresponding to a family of trajectories, which starts off with an exponential expansion phase and then goes over to a matter-dominated era, reaches a maximum expansion and then recontracts. The graph of the radius a against the time t has the peculiar feature that has small amplitude rapid oscillations superimposed on an overall expansion and contraction. These oscillations in radius would cause particle creation of matter fields that were present. Particle creation would dampen the oscillations. The energy of oscillations would go into the particles which lead the universe to expand in the standard big bang manner.

9. The Isotropy of the Universe

In order to investigate the isotropy of the universe, one can extend the isotropic mini-superspace to the Bianchi type IX model with a massive field.[26] The metric is given by

$$ds^2 = -dt^2 + e^{2\alpha(t)}[e^{2\beta}]_{ij}w^i w^j, \tag{9.1}$$

where $\beta_{ij}(t)$ is a 3×3 symmetric traceless matrix and w^i are three 1-forms on the three sphere which obey

$$dw^i = \varepsilon_{ijk}w^i \wedge w^k \tag{9.2}$$

and β can be diagonalized and parametrized as

$$\beta_{ij} = \mathrm{diag}(\beta_+ + \sqrt{3}\beta_-, \beta_+ - \sqrt{3}\beta_-, -2\beta_+). \tag{9.3}$$

Then the Wheeler-DeWitt equation is

$$\left[e^{-p\alpha}\frac{\partial}{\partial\alpha}\left[e^{p\alpha}\frac{\partial}{\partial\alpha}\right] - \frac{\partial^2}{\partial\phi^2} - \frac{\partial^2}{\partial\beta_+^2} - \frac{\partial^2}{\partial\beta_-^2} + e^{6\alpha}m^2\phi^2 + e^{4\alpha}V(\beta_+, \beta_-)\right]\Psi = 0, \tag{9.4}$$

where

$$V(\beta_+, \beta_-) = \frac{1}{2}(e^{-8\beta_+} + e^{4\beta_+ + 4\sqrt{3}\beta_-} + e^{4\beta_+ - 4\sqrt{3}\beta_-})$$

$$- (e^{4\beta_+} + e^{-2\beta_+ + 2\sqrt{3}\beta_-} + e^{-2\beta_+ - 2\sqrt{3}\beta_-}). \tag{9.5}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

As usual, the length scale α is the only time-like coordinate. Its sufficiently negative value corresponds to a very small size of the 3-sphere and when α is sufficiently negative one can ignore the non-derivative potential terms $V \equiv e^{6\alpha}m^2\phi^2 + e^{4\alpha}V(\beta_+, \beta_-)$. Then the wave function has an exponential or power law behavior corresponding to a Euclidean 4-geometry in the classical limit. However, as α increases, the wave function will start to oscillate if V is positive and has a gradient which is "time-like" in configuration space. An oscillating wave function can be interpreted in terms of Lorentzian 4-geometries.

The classical evolutions corresponding to the wave packet is of very small initial velocities $d\beta_+/dt$, $d\beta_-/dt$ and $d\phi/dt$. Lorentzian solutions with these initial conditions will expand exponentially at first because the term $m^2\phi^2$ in the action acts like a cosmological constant. During the exponential expansion β_{ij} will be approximately constant, the scalar factor e^{α} will increase and then contract as in the isotropic massive scalar field model. The scale factor e^{α} will increase by order $\exp(\frac{1}{2}\phi_1^2)$ before ϕ passes through zero, while ϕ_1 is the initial scalar field.

During the period of exponential expansion, the scalar field energy density

$$\rho_m = \left(\frac{d\phi}{dt}\right)^2 + m^2\phi^2 \tag{9.6}$$

will remain constant while the anisotropy energy density

$$\rho_\beta = \left(\frac{d\beta_+}{dt}\right)^2 + \left(\frac{d\beta_-}{dt}\right)^2 + e^{-2\alpha}\left(V + \frac{3}{2}\right). \tag{9.7}$$

will decrease like $e^{-2\alpha}$. In order to resolve the puzzle of the inflationary universe, the period of exponential expansion must be long enough, i.e., $\exp(9\phi_1^2) > 10^{66}$. During the subsequent expansion, the anisotropy energy decreases, generally, faster than the matter-energy density, i.e., at the present day, we have

$$\frac{\rho_\beta}{\rho_m} < \exp(-9\phi_1^2) < 10^{-66}. \tag{9.8}$$

In comparison with the upper limit given by the microwave background, it seems that quantum cosmology can account for the observed isotropy of the universe with a high degree of probability.

10. The Structure of the Universe

The problem of the origin of the structure in the universe has gone beyond the mini-superspace model.[14] The general metric of the 3-geometry with topology S^3 can be written as

$$h_{ij} = a^2(\Omega_{ij} + \varepsilon_{ij}), \tag{10.1}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where Ω_{ij} is the metric on the unit 3-sphere and ε_{ij} is the perturbation superposed on this metric. The perturbation might be expanded in terms of the standard scalar, transverse vector and transverse traceless tensor harmonics. The lapse, shift and the scalar field, the only matter content in the massive scalar model, can also be expanded in terms of the harmonics. By a suitable choice of gauge, one can eliminate all unphysical degrees of freedom, and the only physically meaningful ones are the transverse traceless metric harmonic (graviton) and the scalar field harmonic.

The whole wave function of the universe can be written as a product of that for mini-superspace background and that for each mode of the harmonic.

As shown in Sec. 7 the wave function for the mini-superspace background of the massive scalar model has been treated exactly. Halliwell and Hawking[14] also treat these perturbations to second order. Although the extrinsic time coordinate does not appear in the Wheeler-DeWitt equation, the intrinsic time, i.e., the overall size of the 3-geometry emerges as the time-like coordinate in terms of the supermetric, forces the wave function to evolve. The extrinsic time concept becomes physically meaningful only after we use the WKB ansatz to interpret the wave packet for the mini-superspace background.

To second order the wave function for the coefficient d_n of the harmonic mode n of physical degrees of freedom should satisfy the time dependent Schrödinger equation

$$i\frac{\partial\Psi^{(n)}}{\partial t} = \frac{1}{2}a^{-3}\left[-\frac{\partial^2}{\partial d_n^2} + d_n^2(n^2-1)a^4\right]\Psi^{(n)}. \tag{10.2}$$

For simplicity, we set mass of the scalar field to be zero in (10.2). This equation has the form of the Schrödinger equation for an oscillator with a time dependent frequency $\nu = (n^2-1)^{1/2}a^{-1}$.

The Hartle-Hawking ground state for the perturbation mode is defined by a Euclidean path integral over all regular trajectories, the regularity conditions at the south pole of the distorted S^4 of the mini-superspace background require $d_n = 0$ there. In the Euclidean background one can approximate the harmonic mode as a linear oscillator and the ground state is identified as the minimum excitation. The wave function remains in this state, even in the Lorentzian background, until the wavelength of the mode exceeds the size of the horizon. In the adiabatic approximation the wave function is written as

$$\Psi^{(n)} \simeq \exp\left[-\frac{1}{2}na^2 d_n^2\right]. \tag{10.3}$$

When the wavelength of the mode becomes equal to the horizon size at $a = nH^{-1}$, the wave function will then "freeze." The wave function remains roughly invariant until the mode re-enters the horizon in the matter-dominated era. In terms of the time dependent frequency of the linear oscillator, the quantum state will be a superposition of highly excited states after the re-entry of the horizon.

One would like to compare the result with the observations. The harmonic modes

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

would have led to the temperature anisotropies in the microwave radiation background and to the density perturbations which could account for the origin of galaxies and all other structure in the universe.

For studying the temperature anisotropies, it is more convenient to work in the gauge in which the scalar field is constant on the 3-geometry configurations, on which energy density remains constant and among them there must be a surface of last scattering of the microwave-radiation background. Under this gauge, the scalar and tensor harmonics of metric are the only degrees of freedom, the wave function of the coefficient for the scalar harmonic of the metric satisfies the time dependent equation

$$i\frac{\partial {}^{s}\Psi^{(n)}}{\partial t} \simeq \frac{1}{2}a^{-3}\left[-Y^2\frac{\partial^2}{\partial a_n^2} + a^4(n^2-4)\frac{a_n^2}{Y^2}\right]{}^{s}\Psi^{(n)}, \tag{10.4}$$

where $Y = 3\phi_*$, ϕ_* is the scalar field value at the moment when the mode leaves the horizon.

The harmonic modes a_n and d_n lead to temperature anisotropies of the following form

$$\langle(\Delta T/T)^2\rangle \simeq n^3\cdot(\langle a_n^2\rangle \text{ or } \langle d_n^2\rangle), \tag{10.5}$$

where the factor n^3 is the number of modes with the same scale. From (10.3) and (10.4) one obtains

$$\langle a_n^2\rangle = Y_*^2 n^{-1}a_*^{-2} = Y_*^2 n^{-3}H^2,$$

$$\langle d_n^2\rangle = n^{-1}a_*^{-2} = n^{-3}H^2, \tag{10.6}$$

where a_* is the size of the universe at the moment the wavelength of the mode equals the horizon size. It follows from (10.5) and (10.6) that the temperature anisotropies are

$$\langle(\Delta T/T)^2 \approx Y_*^2 H^2. \tag{10.7}$$

Its observational upper limit of about 10^{-8} restricts, from (10.7), the mass of the scalar field to less than 10^{14} GeV.

If we choose the old gauge, then the density fluctuation at the time that the wavelength comes within the horizon can be obtained as

$$\left\langle\left(\frac{\Delta\rho}{\rho}\right)^2\right\rangle \sim Y^2\frac{\dot\rho_e^2}{\rho_e^2}\frac{\dot a_*^2 a_e^2}{a_*^2\dot a_e^2}, \tag{10.8}$$

where subscript e denotes the moment of re-entry of the horizon. Because Y and $\dot a_*/a_*$ depend only logarithmically on the mode wavelength, this gives an almost scale free spectrum of density fluctuation, i.e., the Harrison-Zeldovich spectrum.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

11. The Time Arrow of the Universe

The universe that we live in certainly does not appear time symmetric. People have realized that there are at least four different "arrows of time" to express the time asymmetry: (a) thermodynamic arrow; (b) electrodynamic arrow; (c) psychological arrow and (d) cosmological arrow. It is generally accepted that the thermodynamic arrow implies the second and third arrows. One would like to know the relation between (a) and (d).[27]

In flat space-time, the thermodynamic arrow is defined by the direction of time in which entropy increases. It has been proved[28] that it is impossible to define the entropy operator in the standard quantum mechanics. The reason is that in quantum mechanics the generator of time translations is identical with the energy operator which is semi-bounded. To overcome this degeneracy, one has to appeal to an enlarged framework, i.e., quantum statistical mechanics, then a quantum state, in general, is expressed by a density matrix, instead of a wave function.

However, the above argument breaks down[27] in the presence of gravity. In quantum cosmology, the time coordinate does not appear explicitly. What appears in the Wheeler-DeWitt equation is the intrinsic time, i.e., the overall size of the universe. In principle, one can solve the scalar constraint (or the Wheeler-DeWitt equation) and formally obtain a Hamiltonian canonically conjugate to the instrinsic time. This Hamiltonian is no longer positive definite. Secondly, if one believes that the quantum state of the universe is described by the Hartle-Hawking state, then there is only one state involved and nothing to do with the ensemble problem, and the time asymmetry may just be due to the special choice of the state, namely, for our case, due to the boundary conditions of the universe. Obviously, the ground state is a pure state. How to incorporate entropy to this state is an important problem. The resolution seems to be that any realistic observation made is only involved with some part of the universe and its interactions with the rest of the universe cause the quantum state of each part to become fuzzy, and describable only by a density matrix, instead of a wave function.

Based on the perturbation calculation of the massive scalar model, Hawking[27] shows that the cosmological arrow defined by the direction of time in which the universe is expanding is related to the thermodynamic arrow of time. This is because in the proposed quantum state the universe would have been smooth and homogeneous when it was small, but disordered and inhomogeneous when it was large. Thus one can make such a conjecture that the thermodynamic arrow would reverse during a contracting phase of the universe or inside black holes.

The arrow of time in the universe is related to the increase of disorder, i.e, the increase of entropy. From the above discussion we know that this arises from the boundary condition of the universe. However, physics laws are symmetric, more accurately, any quantum field theory is invariant under CPT[29] in the absence of gravity. It is known that the sufficient conditions to prove the CPT theorem are: (a) Lorentz invariance; (b) positivity of energy; and (c) local causality. Conditions (a) and (b) are not valid in the presence of gravity and (c) does not make sense when the gravitational field is

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

quantized. Nevertheless, Hawking proves[27] that the universe is invariant under CPT if the quantum state of the universe is defined by the ground state in the Hartle-Hawking sense.

12. The Flatness of the Universe

The Wheeler-DeWitt equation is the quantum counterpart of the scalar constraint. There is some ambiguity in the operator ordering when we replace the momenta in the constraint by the differential operators. To eliminate the ambiguity, Hawking and Page[15] present an ansatz, that is, the equation takes a form which is invariant under coordinate transformations in generalized superspace with the metric

$$ds^2 = \iint N(x)^{-1}\left[\frac{1}{32\pi} G^{ijkl}(x,x')\gamma_{ij}(x)\gamma_{kl}(x') + h^{1/2}\delta^3(x-x')\mu(x)\mu(x')\right] d^3x\, d^3x', \tag{12.1}$$

where, for simplicity, we have assumed that the scalar field is the only matter content, and we use $\gamma_{ij}(x)$ and $\mu(x)$ to denote a symmetric tensor field and a scalar field on the generalized superspace, and

$$G^{ijkl}(x,x') = \frac{1}{2}h^{1/2}[h^{ik}(x)h^{jl}(x') + h^{il}(x)h^{jk}(x') - 2h^{ij}(x)h^{kl}(x')]\delta^3(x-x'). \tag{12.2}$$

Under the Hawking-Page ansatz, the most natural form of the Wheeler-DeWitt equation is written as

$$\left(-\frac{1}{2}\nabla^2 + V\right)\Psi(h_{ij},\phi) = 0, \tag{12.3}$$

where ∇^2 is the Laplacian in the metric (12.1) and the potential V is

$$V = \int Nh^{1/2}\left[\frac{1}{16\pi}(-{}^3R + 2\Lambda) + U(\phi)\right] d^3x, \tag{12.4}$$

where $U(\phi)$ is the potential of the scalar field ϕ. In the Hawking massive scalar model[23] it has been specified that $U(\phi) = \frac{1}{2}m^2\phi^2$. However, in order to study the flatness problem, we shall not restrict it to be of this form.

The Wheeler-DeWitt equation takes the form of the Klein-Gorden equation. Thus the current

$$J = \frac{1}{2}(\Psi\nabla\bar{\Psi} - \bar{\Psi}\nabla\Psi) \tag{12.5}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

is conserved. One is tempted to define its time-like component (with respect to the metric (12.1)) as probability. It turns out that this is impossible. Firstly, it is not positive definite. Secondly, since the wave function is real for our case, J must vanish. Instead Hawking defines the probability of finding a 3-surface S with metric h_{ij} and scalar ϕ to be proportional to $|\Psi|^2$. If one chooses N in (12.1) to be a function of x only and independent of h_{ij}, the probability is uniquely defined under metric (12.1). In the wave packet region, one can associate a probability to each component of the propagating wave $C_n e^{iS_n}$. The phase S_n satisfies the Hamilton-Jacobi equation

$$\frac{1}{2}(\nabla S_n)^2 + V = 0; \tag{12.6}$$

then C_n must obey

$$\frac{1}{2}\nabla^2 C_n + i(\nabla S_n)\cdot(\nabla C_n) + \frac{1}{2}iC_n\nabla^2 S_n = 0, \tag{12.7}$$

where the last term of (12.7) is the quantum correction and can be neglected in the WKB approximation. Equation (12.7) becomes the conservation equation for the current

$$J_n = |C_n|^2 \nabla S_n . \tag{12.8}$$

The probability in a region A is proportional to the product of the coordinate time Δt that the classical solutions spend in A and the flux of J_t across it.

As an application of the probability definition Hawking and Page compute the probability of the density parameter

$$\Omega = \frac{\rho}{\rho - \dfrac{k}{a^2}} \tag{12.9}$$

in a mini-superspace model. The 3-geometry is restricted to be locally homogeneous and isotropic with $k = +1, 0, -1$ corresponding to 3-sphere, flat 3-space and 3-hyperboloid factored by a discrete group.

The Wheeler-DeWitt equation is

$$\left[a\frac{\partial}{\partial a}a\frac{\partial}{\partial a} - \frac{\partial^2}{\partial p^2} - ka^4 + 2a^6 U(\phi)\right]\Psi(a,\phi) = 0, \tag{12.10}$$

where a is the scale length of the 3-geometry. It is assumed that $U(\phi)$ decreases monotonically from a large value of ϕ to a minimum of zero at some ϕ which one may

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

shift to be at $\phi = 0$. Thus the qualitative behavior of the classical trajectories in the wave packet region is similar to that of the Hawking model.

The quantity

$$E = a^3\left[\left(\frac{d\phi}{dt}\right)^2 + 2U\right], \tag{12.11}$$

which may be interpreted as the energy of the scalar field within the whole universe, is approximately conserved along each classical trajectory at the matter-dominated era. In the inflationary era, E grows rapidly and quantity (12.11) can be approximately given by

$$E(\phi_0) \simeq K\exp\left[18\int_0^{\phi_0}\left[\frac{dU}{d\phi}\right]^{-1}U(\phi)\,d\phi\right], \quad (K \text{ const.}), \tag{12.12}$$

where each trajectory is denoted by ϕ_0, the value of ϕ where the wave function first crosses zero as a is increased from zero. Page[20] shows that the probability flux per range ϕ_0 is independent of ϕ_0, thus the probability per proper time contributed by the various classical trajectories is dominated by those coming from infinitely large ϕ_0.

They also calculate the unconditional probability distribution of the density parameter Ω,

$$\rho(\Omega)\,d\Omega = \left[\frac{\Omega - 1}{K^3}\right]^{1/2}\frac{d\Omega}{\Omega^2}. \tag{12.13}$$

One does not get a sharp prediction for $\Omega = 1$. The physical relevant quantity is the probability distribution for Ω within some fixed interval of density $d\rho$, for which the existence of an observer is possible. Because the contribution to the probability is proportional to the proper time and the probability flux is independent of ϕ_0 for sufficiently large $|\phi_0|$, one can show from (12.9) that the probability distribution will be dominated by trajectories from infinitely large $|\phi_0|$ which will have $\Omega = 1$ when their density lies between ρ and $\rho - d\rho$. This result crucially depends on the choice of metric of (12.1).

13. The Dimension of the Universe

The Kaluza-Klein theory[30] was the first attempt to unify gravitation with the electromagnetic interaction—the only alternative interaction known in the twenties. Now $N = 8$ supergravity theory[31] provides a more realistic framework for the unification of all four interactions, and it can be derived in terms of a simply theory in a space-time of higher dimensionality, i.e., simple supergravity in eleven dimensions. It is believed that the extra seven dimensions are perhaps realistic and are reduced to the unobserved Planck length scale by the so-called spontaneous compactification. The

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

symmetry of this internal space corresponds to the internal symmetries of low energy physics. Since its relevant scale is inversely correlated with the gauge coupling constant in the theory,[32] it is of great interest to investigate its time evolution.

For a 5-dimensional $R \times S^3 \times S^1$ model with a positive cosmological constant, as in the de Sitter model, one can get the wave function analytically at the semi-classical approximation[11,33]

$$\Psi[a,b] \simeq N_0 \exp\left(\frac{1}{4}a^2 b\sqrt{1-H^2a^2}\right), \quad a < H^{-1},$$

$$\Psi[a,b] \simeq N_0' \cos\left(\frac{1}{4}a^2 b\sqrt{H^2a^2-1}+\delta\right), \quad a > H^{-1},$$

$$H = \sqrt{\frac{\Lambda}{6}}, \quad \delta = \text{const.}, \tag{13.1}$$

where a is the length scale of the three sphere S^3 and b is the radius of the circle S^1 of the fifth dimension.

The classical trajectories corresponding to the wave packet are

$$ds^2 = \left(\frac{1}{6\pi^2}\right)^{2/3}\left[-dt^2 + \frac{6}{\Lambda}\cosh^2\left(\sqrt{\frac{\Lambda}{6}}t\right)d\Omega_3^2 + c_3^2\sinh^2\left(\sqrt{\frac{\Lambda}{6}}t\right)dw^2\right],$$

$$c_3 = \text{const.}, \quad 0 \le w < 2\pi. \tag{13.2}$$

Thus the length scales of S^3 and S^1 will eventually exponentially increase at the same rate, i.e., the spontaneous compactification will not occur. One can get compactification by using an anisotropic cosmological constant in the model with the Euclidean action

$$\bar{I} = \frac{1}{2}\int\left\{\frac{a^3 b}{12}\left[\frac{3\dot{a}^2}{a^2}+\frac{\dot{b}^2}{b^2}-\left(\frac{3\dot{a}}{a}+\frac{\dot{b}}{b}\right)^2-\frac{b}{a^2}\right]+\frac{a^{7/3}b^{5/3}}{b}\Lambda\right\}, \tag{13.3}$$

where a dot denotes the derivative in terms of τ.

For the higher dimensional model $R \times S^3 \times S^n$ $(n \ge 2)$, the metric takes the form[33]

$$ds^2 = -dt^2 + a^2(t)\,d\Omega_3^2 + b^2(t)\,d\Omega_n^2. \tag{13.4}$$

By introducing the scale and the anisotropy parameters

$$a = R\gamma^n, \quad b = R\gamma^{-3}, \tag{13.5}$$

the Wheeler-DeWitt equation is written as

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$\left[-\frac{1}{R^p}\frac{\partial}{\partial R}R^p\frac{\partial}{\partial R}+\frac{\partial^2}{R^2\partial\rho^2}+R^{2n+2}(6\cdot e^{2n\rho\sqrt{(n+2)/3n}}\right.$$

$$\left.+\,n(n-1)e^{6\rho\sqrt{(n+2)/3n}}-2\Lambda R^2)\right]\Psi(R,\rho)=0,$$

$$\rho\equiv\sqrt{\frac{3n}{n+2}}\ln\gamma. \tag{13.6}$$

As in the Hawking massive scalar model, the null cone $R=0$ is the boundary of the configuration space. The regularity conditions at the south pole of all Euclidean manifolds restrict $\Psi=$ const. there. This boundary condition is sufficient to evolve (13.6) for the whole configuration space, i.e., for the future of the null cone. Near the null cone, the wave function behaves exponential, and it will start to oscillate at some space-like segment of the zero-potential surface

$$\frac{6}{a^2}+\frac{n(n-1)}{b^2}-2\Lambda=0,$$

$$\frac{3n-\sqrt{3n^2+6n}}{6}<\left(\frac{b}{a}\right)^2<\frac{3n+\sqrt{3n^2+6n}}{6}. \tag{13.7}$$

Thus the wave function represents an ensemble of classical trajectories which satisfy the Hamilton-Jacobi equation. As usual, the ground state proposal imposes a necessary and sufficient restriction so that there is one and only one trajectory through each point in the wave packet region. All classical orbits represented by the wave packet begin with a zero initial velocity from segment (13.7).

It is reasonable to assume that the universe we live in is of the greatest probability. The relative probability is evaluated by a path integral

$$\Psi^*\Psi=\int_{c+c^*}\delta[g_{\mu\nu}]\delta[\phi]\exp(-\bar{I}[g_{\mu\nu},\phi]). \tag{13.8}$$

The path integral is summed over all compact positive definite $(n+4)$-metrics which have the configuration $S^3(a)\times S^n(b)$ as a section. One can view the transition to the wave packet region as a quantum tunneling from Euclidean compact $(n+4)$-geometries. A gravitational instanton is defined to be a complexified solution of classical field equations with some real compact sector. Therefore, it dominates the contribution to the path integral (13.8). It follows that the most probable evolution of the universe must be quantum mechanically penetrated through the zero potential surface from an instanton. For our trial model, the only instantons are $S^4\times S^n$ and $S^3\times S^{n+1}$. Turning to the Lorentzian region, it follows that the universe most probably

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

evolves with an exponentially expanding external space and a static internal space. However, since the dimension of the metric form is put in by hand, the dimension problem remains unsolved. One must appeal to more realistic models.

The Kaluza-Klein theory with $d = 11$ simple supergravity is a more realistic model. Supersymmetry compels us to introduce a spin-$\frac{3}{2}$ gravitino field and a 3-index anti-symmetric tensor A_{MNP} into the theory. For simplicity, we assume the gravitino field to be zero. Then the field equations are[30]

$$R_{MN} - \frac{1}{2} g_{MN} R = \frac{1}{48}(8F_{MPQR}F_N^{PQR} - g_{MN}F_{SPQR}F^{SPQR}), \tag{13.9}$$

$$F^{MNPQ}_{;M} = \left[\frac{-\sqrt{2}}{2\cdot(4!)^2}\right]\cdot \eta^{M_1\ldots M_8 NPQ}F_{M_1\ldots M_4}F_{M_5\ldots M_8}, \quad (0 \le M, N, P, Q, R, S, \ldots \le 10), \tag{13.10}$$

where $F_{MNPQ} = 4!\,\partial_{[M}A_{NPQ]}$ and $\eta^{A\ldots N} = |g|^{-1/2}\varepsilon^{A\ldots N}$. In the following we shall work with a mini-superspace model with only a finite number degrees of freedom: the $d = 11$ space-time takes a product form of Einstein spaces $M^m \times M^n$ $(m + n = 11)$, and all F_{FNPQ} components with mixed indices m and n are assumed to vanish. Now we attempt to find the wave function of ground state, specifying M^m to be $R \times S^{m-1}$ and M^n to be S^n, where R denotes time, S^{m-1} and S^n are interpreted as the external and internal spaces, respectively. The Wheeler-DeWitt equation is hyperbolic with the geometric average of the scale length γ as the only time-like coordinate. In configuration space, when γ is sufficiently small, the action for Euclidean metrics is very small. The wave function takes some polynomial and exponential forms there. One expects that as soon as it reaches a space-like segment of the zero-potential surface, the wave function would start to oscillate and the universe would begin to evolve. Unfortunately, the zero-potential surface remains time-like everywhere and no Lorentzian region exists. This corresponds to the fact that there does not exist any classical Lorentzian solution under this ansatz.

However, if we assume $M^m = R \times H^{m-1}$ (H^{m-1} is an $(m-1)$-dimensional hyperboloid), the situation changes drastically. One can find classical solutions of metric signature $(\underbrace{-\,-\cdots-}_{m}, \underbrace{+\,+\cdots+}_{n})$ with $S^{m-1} \times S^n$ as the configuration (S^{m-1} is considered as a compact real sector of complexified H^{m-1}). These solutions dominate the path integral for the ground state. Returning to the H^{m-1} configuration, the wave function oscillates and represents the classical orbits of ground state. Possible orbits exist for $m = 2, 3, \ldots, 9$, which means one can have microscopic space-time of 2, 3, ..., 9 dimensions. Considering dimensionality, one hopes that the most probable evolution of the universe must have $m = 4$. This is indeed the case.

In order to obtain the most probable evolution, one needs to find the instanton solution. In the subspace S^m $(m = 1, 2, 3)$ F vanishes and for our ansatz the right-hand

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

side of Eq. (13.10) for S^{11-m} vanishes. It follows that F must be a harmonic in S^{11-m}. However, we learn from de Rham cohomology[34] that $H^4(S^4) = 1$ and $H^4(S^m) = 0$ $(m \neq 4)$. So there is no nontrivial instanton. If $m = 5$, both F components in S^5 and S^6 must be harmonics and so vanish. By the dimension duality, there does not exist a nontrivial instanton for $m = 10, 9, 8, 6$. The only nontrivial instanton is of $m = 4$ or $m = 7$ and the F components in S^7 does not need to be a harmonic. Fortunately, some instantons have been found. Their Lorentzian versions are Freund-Rubin,[35] Duff-Pope,[36] Englert,[37] and Englert-Rooman-Spindel[38] space. These spaces are products of a 4-dimensional anti-de Sitter space and a round or squashed 7-sphere. Perhaps, these cases exhaust all instanton solutions with this ansatz. Note that, de Sitter space and anti-de Sitter space are two real sectors of the 4-sphere. Before the discovery of quantum cosmology, one did not know how to exclude the cases $m = 5, 6$.

It is believed that this 4-dimensional anti-de Sitter space is our macroscopic space-time. The static internal space represents the internal symmetry of particle physics. The gauge coupling constant inversely related to the size of the internal space S^7 must be time independent. The 7-dimensional sector of Freund-Rubin space can be replaced by products $S^2 \times S^2 \times S^3$, $S^2 \times S^5$ or $S^3 \times S^4$. In any case, one cannot avoid a 4-dimensional sector. A conjecture is as follows: for the most general ansatz any instanton must have a 4-dimensional sector. This may be the main reason why space-time is apparently 4-dimensional.

14. Further Developments

In principle one can predict everything in the universe solely from physical laws within the framework of quantum cosmology. However, in practice the calculation is very difficult.

Some effort has been made to investigate the quantum creation of primordial black holes in quantum cosmology.[39] It is shown that a black hole of considerable size cannot form in the beginning of the universe.

In discussing the problem of time arrows, Hawking presents a conjecture[36] that the time arrow would reverse inside black holes. The wave function of a mini-superspace universe with a 3-surface of topology $S^1 \times S^2$ and a massive scalar field describes the quantum state of a black hole interior. However, since the solution is not time symmetric around the moment of maximum expansion, entropy will not start decreasing there. One cannot use this model to confirm Hawking's conjecture.[40]

Recently, Halliwell[41] presented a model in 6-dimensional Einstein-Maxwell theory. In the classical limit, the wave packet corresponds to a superposition of classical solutions in which two of the dimensions remain small while the other four expand exponentially.

Another way out of the problem on a non-positive definite Euclidean action (5.2) is to add quadratic curvature terms in the action.[42] In this case the ground state wave function still describes an inflationary universe.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

It might be interesting to discuss a model with fermion fields as the matter content. In the Planck time, the quantum state of the universe might be supersymmetric, so it is very attractive to study supersymmetric quantum cosmological models.

References

1. S. Weinberg, *Gravitation and Cosmology* (John Wiley, New York, 1972).
2. W. Rindler, *Mon. Not. R. Astron. Soc.* **116** (1956) 663.
3. R. H. Dicke and P. J. E. Peebles, in *General Relativity, An Einstein Centenary Survey*, eds. S. W. Hawking and W. Israel (Cambridge Univ. Press, Cambridge, 1979).
4. T. W. Kibble, *J. Phys.* **A9** (1976) 1387.
5. A. A. Starobinsky, *Phys. Lett.* **91B** (1980) 99; K. Sato, *Phys. Lett.* **99B** (1981) 66; A. H. Guth, *Phys. Rev.* **D23** (1981) 347; A. D. Linde, *Phys. Lett.* **108B** (1982) 389; S. W. Hawking and J. G. Moss, *Phys. Lett.* **110B** (1982) 35; A. Albrecht and P. J. Steinhardt, *Phys. Rev. Lett.* **48** (1982) 1220.
6. S. Coleman, *Phys. Rev.* **D15** (1977) 2929; S. Coleman and F. De Luccia, *Phys. Rev.* **D21** (1980) 3305.
7. S. W. Hawking, I. G. Moss and J. M. Stewart, *Phys. Rev.* **D26** (1982) 2681; Wu Zhong Chao, *Phys. Rev.* **D28** (1983) 1898.
8. S. Coleman and E. Weinberg, *Phys. Rev.* **D7** (1973) 1888.
9. J. D. Barrow, in *The Very Early Universe*, eds. G. W. Gibbons, S. W. Hawking and S. T. C. Siklos (Cambridge Univ. Press, Cambridge, 1983); W. Boucher and G. W. Gibbons, *ibid.*
10. S. W. Hawking, *Phys. Lett.* **115B** (1982) 295; A. H. Guth and So-Young Pi, *Phys. Rev. Lett.* **49** (1982) 1110.
11. Wu Zhong Chao, *Phys. Rev.* **D30** (1984) 286.
12. Ya. B. Zeldovich, *Mot. Not. R. Astron. Soc.* **160** (1972) 1; E. R. Harrison, *Phys. Rev.* **D1** (1970) 2726.
13. B. S. DeWitt, *Phys. Rev.* **160** (1968) 1113; J. A. Wheeler, in *Battelle Rencontres*, eds. C. M. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).
14. J. J. Halliwell and S. W. Hawking, *Phys. Rev.* **D31** (1985) 1779.
15. S. W. Hawking and D. N. Page, *Nucl. Phys.* **B264** (1986) 185.
16. S. W. Hawking, in *General Relativity: An Einstein Centenary Survey*, eds. S. W. Hawking and W. Israel (Cambridge Univ. Press, Cambridge, 1979).
17. S. W. Hawking, in "Astrophysical cosmology," *Pontifical Academiae Scientarum Varia* **48** (1982) 563; J. B. Hartle and S. W. Hawking, *Phys. Rev.* **D28** (1983) 2960; S. W. Hawking, in *Relativity, Groups and Topology* II, eds. B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).
18. G. W. Gibbons, S. W. Hawking and M. J. Perry, *Nucl. Phys.* **B138** (1978) 141.
19. J. B. Hartle and K. Schleich, Santa Barbara preprint TH-70, 1985.
20. D. N. Page, in *Quantum Concepts in Space and Time*, eds. C. J. Isham and R. Penrose (Oxford Univ. Press, London, 1985).
21. Wu Zhong Chao, *Phys. Rev.* **D31** (1985) 3079.
22. J. B. Hartle, *Phys. Rev.* **D29** (1984) 2730.
23. S. W. Hawking, *Nucl. Phys.* **B239** (1984) 257; S. W. Hawking and Z. C. Wu, *Phys. Lett.* **151B** (1985) 15.
24. I. G. Moss and W. A. Wright, *Phys. Rev.* **D29** (1984) 1067.
25. S. W. Hawking and J. Luttrell, *Nucl. Phys.* **B247** (1984) 250.
26. S. W. Hawking and J. Luttrell, *Phys. Lett.* **143B** (1984) 83.
27. S. W. Hawking, *Phys. Rev.* **D32** (1985) 2489; D. N. Page, *Phys. Rev.* **D32** (1985) 2496.
28. I. Prigogine, *From Being to Becoming* (W. H. Freeman & Co., San Francisco, 1980).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

29. R. F. Streater and A. S. Wightman, *PCT, Spin, Statistics and All That* (Benjamin, New York, 1964).
30. Th. Kaluza, *Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl.* 966 (1921). O. Klein, *Z. Phys.* **37** (1926) 895.
31. Cremmer, B. Julia and J. Scherk, *Phys. Lett.* **76B** (1978) 409.
32. A. Salam and J. Strathdee, *Ann. Phys.* (NY) **141** (1982) 316.
33. Wu Zhong Chao, *Phys. Lett.* **146B** (1984) 307; Wu Zhong Chao and Hu Xiao Ming, *Phys. Lett.* **149B** (1984) 87, **155B** (1985) 237; Wu Zhong Chao, *Gen. Rel. Grav.* **17** (1985) 1217.
34. T. Eguchi, P. B. Gilkey and A. J. Hanson, *Phys. Reports* **66** (1980) 213.
35. G. O. Freund and M. A. Rubin, *Phys. Lett.* **97B** (1980) 233.
36. M. J. Duff and C. N. Pope, in *Supersymmetry and Supergravity*, eds. S. Ferrara, J. G. Taylor and P. van Nieuwenhuizen (World Scientific, Singapore, 1982).
37. F. Englert, *Phys. Lett.* **119B** (1982) 339.
38. F. Englery, M. Rooman and P. Spindel, *Phys. Lett.* **127B** (1983) 47.
39. L. Z. Fang and M. Li, *Phys. Lett.* **169B** (1986) 28.
40. R. Laflamme, DAMPT preprint, 1985.
41. J. J. Halliwell, *Nucl. Phys.* **266** (1986) 228.
42. G. T. Horowitz, *Phys. Rev.* **D31** (1985) 1169.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Copyright © 1968 by W. A. Benjamin, Inc.

IX

Superspace and the Nature of Quantum Geometrodynamics

JOHN ARCHIBALD WHEELER

Allowable History Selected out of Arena of Dynamics by Constructive Interference *242*
Superspace *246*
Electricity as Lines of Force Trapped in the Topology of Space *263*
The Energy of the Vacuum *269*
A Particle as a Geometrodynamical Exciton *270*

Problem 1. "Derivation" of "Einstein–Hamilton–Jacobi Equation" *273*
Problem 2. Structure of Superspace *278*
Problem 3. Initial Conditions *295*
References and Notes *295*

ALLOWABLE HISTORY SELECTED OUT OF ARENA OF DYNAMICS BY CONSTRUCTIVE INTERFERENCE

Particle dynamics takes place in the arena of spacetime. Geometrodynamics takes place in the arena of superspace. One needs only to mount to this point of view to have the whole content of Einstein's theory spread out before his eyes, as in the outlook from a mountain peak, and as well quantum geometrodynamics as classical geometrodynamics. What is the nature of the landscape that we see from this height? What structures can we hope to build upon this landscape? And what kinds of mysteries are hidden in the mists beyond?

In looking at the terrain of dynamics, happily everyone by now is long accustomed to the Hamilton–Jacobi theory. It belongs entirely to the world

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

of classical physics. Yet it carries one in an instant into the world of the quantum of action.

When better has one ever seen the transition from quantum to classical than in the theory of a particle in motion? Call the potential $V = V(x)$. Call the energy of the particle E. Then there is not the slightest hope of discussing the movement of the particle in space and time. Complementarity forbids! The wave function is spread out all over space. That one sees in no way more easily than through the semiclassical approximation for the probability amplitude function,

$$\psi_E(x, t) = \begin{pmatrix}\text{SLOWLY VARYING} \\ \text{AMPLITUDE FUNCTION}\end{pmatrix} \exp\left(\frac{i}{h}\right) S_E(x, t) \tag{1}$$

It is of no help in localizing the probability distribution that the Hamilton–Jacobi function S has in many applications a value large in comparison with the quantum of angular momentum $h = 1.02 \times 10^{-27}$ g-cm^2/sec. It is of no help that this "dynamical phase"—to give S another name—obeys the simple Hamilton–Jacobi law of propagation,

$$\frac{\partial S}{\partial t} = H\left(\frac{\partial S}{\partial x}, x\right)$$

$$= \left(\frac{1}{2m}\right)\left(\frac{\partial S}{\partial x}\right)^2 + V(x) \tag{2}$$

And finally, it is of no help that the solution of this equation for a particle of energy E is extraordinarily simple,

$$S(x, t) = -Et + \int^x \{2m[E - V(x)]\}^{1/2}\, dx + \delta_E \tag{3}$$

The probability is still spread all over everywhere! There is not the slightest trace of anything like a localized world line, $x = x(t)$!

How old the idea of building wave packets out of monofrequency waves—and how easy! The probability amplitude is now a superposition of terms, qualitatively of the form

$$\psi(x, t) = \psi_E(x, t) + \psi_{E+\Delta E}(x, t) + \cdots \tag{4}$$

Destructive interference takes place almost everywhere. The wave packet is concentrated in the region of constructive interference. There the phases of the various waves agree; thus

$$S_E(x, t) = S_{E+\Delta E}(x, t) \tag{5}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

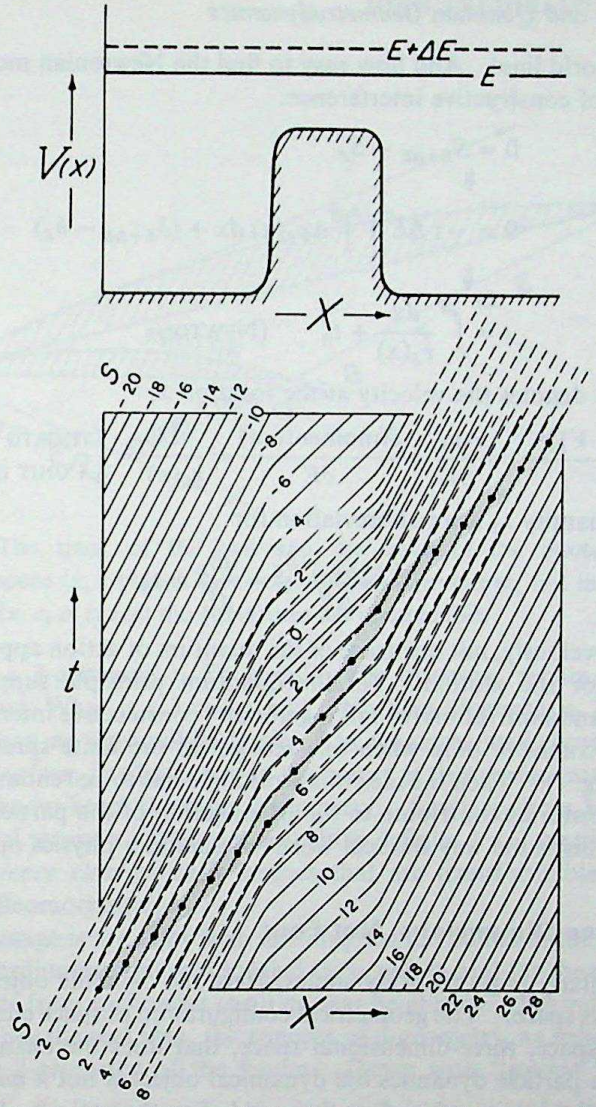


FIGURE 1. "Motion" and "world line" of a particle appear in quantum mechanics as the consequence of interference between wave trains that extend over all space. Above, potential energy as a function of distance for a model problem.

Below, smooth lines numbered -20, -18, ..., 28 are wave crests of probability amplitude function $\psi_E(x, t) \sim$ (slowly varying amplitude factor) $\exp(i/\hbar)S(x, t)$ for energy E. Dashed lines, same for energy $E + \Delta E$. Shaded area, region of constructive interference ("wave packet"). Black dots mark locus of classical world line ($S_{E+\Delta E} = S_E$).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

At last a world line! And how easy to find the Newtonian motion from this condition of constructive interference:

$$0 = S_{E+\Delta E} - S_E$$
$$\Downarrow$$
$$0 = -t\,\Delta E + \int^x \Delta p_E(x)\,dx + (\delta_{E+\Delta E} - \delta_E)$$
$$\Downarrow$$
$$t = \int^x \frac{dx}{v_E(x)} + t_0 \qquad \text{(NEWTON)} \tag{6}$$

Here $v_E(x)$ denotes the velocity at the location x,

$$\frac{\Delta[2m(E-V)]^{1/2}}{\Delta E} = \frac{\Delta p_E}{\Delta E} \to \frac{\partial(\text{momentum})}{\partial E} = \frac{1}{v_E(x)} = \begin{pmatrix}\text{TIME TO COVER}\\ \text{A UNIT DISTANCE}\end{pmatrix} \tag{7}$$

and the quantity t_0 is an abbreviation for

$$\frac{\delta_{E+\Delta E} - \delta_E}{\Delta E} \to \frac{d\delta_E}{dE} \equiv t_0 \tag{8}$$

Marvelously, not one trace of the quantum of action appears in the final solution for the motion. Yet the quantum principle supplies the whole rationale and motivation for talking about "constructive interference." The quantum comes in only when one recognizes the finite spread of the wave packet (Fig. 1). Then the idea of a world line has to be renounced. A whole range of histories contribute to the propagation of the particle from start to finish. This is the way the real world of quantum physics operates!

Three Dimensions, Not Four

Similarly in geometrodynamics: Here the dynamic object is not spacetime. It is space. The geometrical configuration of space changes with time. But it is space, three-dimensional space, that does the changing. No surprise! In particle dynamics the dynamical object is not x *and* t, but only x. How to tell this to our friends in the world of mathematics? For so long they have heard us say that it was in default of the fourth dimension that Riemann could not have discovered general relativity. First there had to come special relativity and spacetime and the fourth dimension. Otherwise how could one have had any possibility to connect gravitation with the curvature of spacetime (Fig. 2)? This understood, how can physicists change their minds and "take back" one dimension? The answer is simple. A decade and more of work by Dirac, Bergmann, Schild, Pirani, Anderson, Higgs, Arnowitt, Deser, Misner, DeWitt, and others has taught us through many a hard knock that Einstein's geometrodynamics deals with the dynamics of geometry: of 3-geometry, not 4-geometry [1, 2].

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

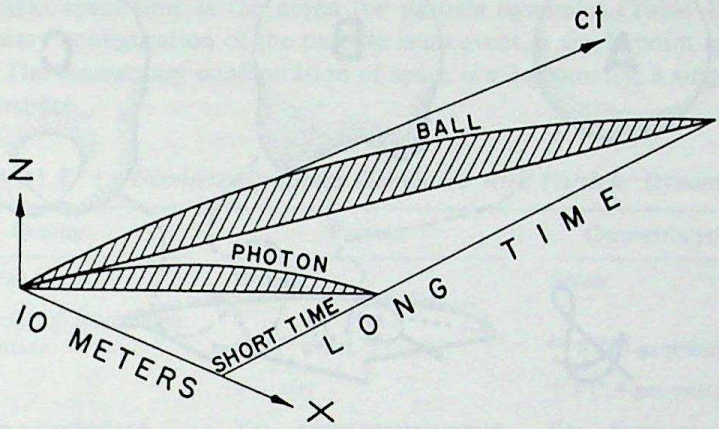


FIGURE 2. The track of the ball and the track of the photon through space (x, z plane) have very different curvatures, but in space time (x, z, ct space) the curvatures are comparable.

What is a 3-geometry? To simplify the question, rephrase it in an everyday context: What is a 2-geometry? Nothing illustrates a 2-geometry more clearly than an automobile fender. In whatever way coordinates are painted on its surface, in whatever way the points of that surface are named or renamed, the fender keeps the same 2-geometry. Similarly for a 3-geometry. In mathematical terms, a ${}^{(3)}\mathcal{G}$ is not a positive definite 3×3 metric; instead, it is an *equivalence class* of such metrics that are transformable, one into another, by diffeomorphisms.

Mere baggage is the right term not only for the points of the space, but also for the coordinates employed to label these points, and for the metric that tells the distance from each point to all its near neighbors. Behind all of that paraphernalia lies the real idea, the concept of 3-geometry, as solid and substantial as the 2-geometry of the fender. Down with "points"; up with "geometry"!

SUPERSPACE

One climbs up to the concept of geometry only to find a new height beyond—superspace [3]. Superspace is the manifold, a single point of which stands for an entire 3-geometry (Fig. 3). A 3-geometry stands at the halfway mark between point and superspace: risen though it is above the concept of point by abstractification, any one geometry of space counts as only a single point of superspace.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

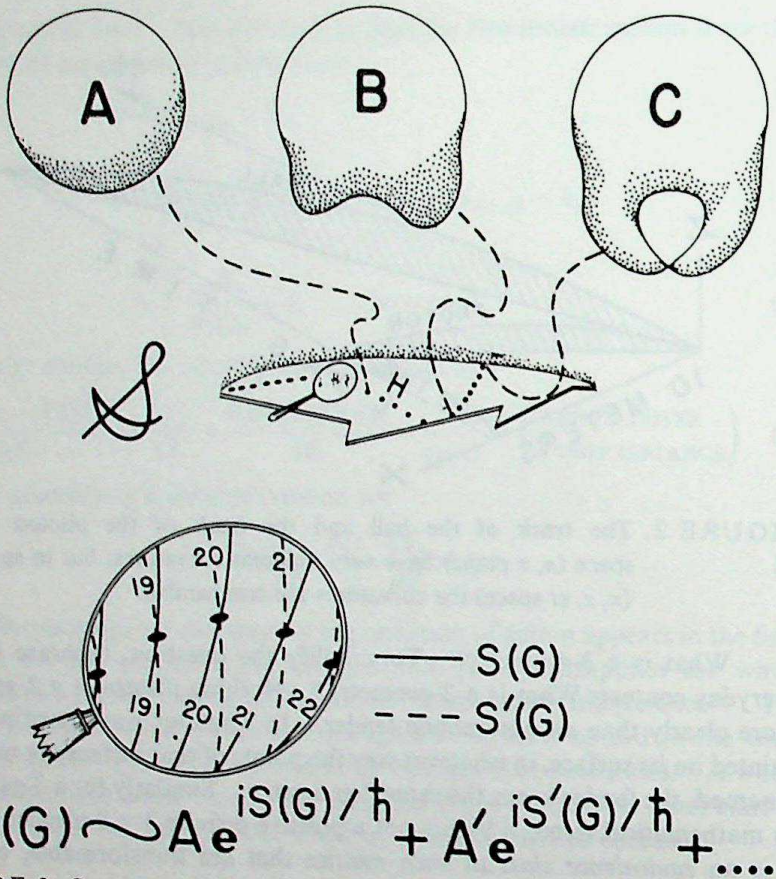


FIGURE 3. Superspace $\mathcal{S}$ is the manifold each of whose "points" A, B, C, ... is an abbreviation for one 3-geometry. A submanifold H of $\mathcal{S}$ is the "classical history of the geometry of space" when space has been started off under some particular set of dynamical initial conditions. In other words, H consists of all those spacelike 3-geometries that can be obtained as spacelike sections through one particular 4-geometry (that satisfies Einstein's classical field equations). The 3-geometries of the classical history H may alternatively but equivalently be distinguished from other 3-geometries by the fact that they satisfy the classical "condition of constructive interference" $S({}^{(3)}\mathcal{G}) = S'({}^{(3)}\mathcal{G}) = S''({}^{(3)}\mathcal{G}) = \cdots$ ("coincidence of wave crests" in lower magnified view of a region of superspace). In quantum theory the wave packet is not localized with unlimited sharpness. The probability amplitude ψ has significant values for 3-geometries at some small "distance" in superspace on each "side" of the classical history H; hence the "quantum fluctuations in the geometry of space" (depicted symbolically in more detail in Fig. 5).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Superspace is the arena for geometrodynamics, just as Lorentz–Minkowski space-time is the arena for particle dynamics (Table 1). The momentary configuration of the particle is an event, a single point in space-time. The momentary configuration of space is a 3-geometry, a single point in superspace.

TABLE 1 *Geometrodynamics Compared with Particle Dynamics*

Quality	Particle	Geometrodynamics
Dynamical entity	Particle	Space
Descriptors of momentary configuration	x, t ("event")	${}^{(3)}\mathcal{G}$ ("3-geometry")
History	$x = x(t)$	${}^{(4)}\mathcal{G}$ ("4-geometry")
History is a stockpile of configurations?	Yes. Every point on world line gives a momentary configuration of particle	Yes. Every spacelike slice through ${}^{(4)}\mathcal{G}$ gives a momentary configuration of space
Dynamic arena	Spacetime (totality of all points x, t)	Superspace (totality of all ${}^{(3)}\mathcal{G}$'s)

Is superspace a proper manifold? Its construction is simple enough: call a 3-geometry a "point," and put all such points together. Or why limit attention to 3-geometries with positive definite metric? Build "extended superspace"; call every 3-geometry a "point" even if its signature is not $+++$, and put all these "points" together. Does the resulting mathematical object possess a reasonable topology? Is it a true manifold? No! In a proper manifold each point has a neighborhood homeomorphic to an open set in a Banach space, and two distinct points have disjoint neighborhoods. Not so here. In an important investigation Michael Stern has shown [4] that for each point of extended superspace there exists another point, distinct from it, which nevertheless cannot be isolated from it by open sets. The topology is not Hausdorf. Extended superspace is not a manifold. It can hardly be regarded as an acceptable arena for dynamics.

Superspace, in contrast, Stern has shown, does have Hausdorf topology and does constitute a manifold. According to these mathematical considerations, *superspace is the proper arena for geometrodynamics.*

Physical considerations point to the same conclusion. To insist that a 3-geometry shall have positive definite metric is to guarantee that no light ray can traverse this 3-geometry. No physical effect can propagate from one point of the ${}^{(3)}\mathcal{G}$ to another. A physical quantity local to the one point and a physical quantity local to the other have zero reciprocal coupling. They commute. Such quantities lend themselves to simultaneous specification over

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the entire 3-geometry. No simpler example exists of a "complete observation" in quantum geometrodynamics. Closely related is the initial value problem of classical geometrodynamics.

"Observation?" Does not observation imply an observer? If there is an observer, does he not respond to every event on his past light cone? And consequently does not that past light cone with its $+ + 0$ metric supply the appropriate geometry on which to specify physical conditions? No, and for two reasons. First, to know the state of the geometry on a past light cone, however completely, is still to be powerless to predict from Einstein's field equations the future of the geometry. Predict the 4-geometry within the past light cone? Yes [5]. Outside—to the future? No. Into the domain of the future, influences flow from afar without ever once impinging upon the light cone. "Demidynamics"—prediction into the past—one can do; full dynamics, no. The fault is in the choice of initial value hypersurface. The cone creates an unsymmetrical divide between past and future. How different from a spacelike initial value hypersurface! The initial value data on it allow a prediction of the complete geometrodynamical history, past and future.

Second, the original argument was mistaken that one should consider "a light cone converging upon an observer." The "observer" of dynamical theory is not and cannot be a single detector of events either in special relativity or in general relativity. Instead he "collects the printout" [6] from a multitude of detectors dotted densely about. Each of his detectors is sensitive only for an instant. These instants of sensitivity bear a spacelike relationship each to the other [7]. Collectively they define a spacelike hypersurface, a "simultaneity," a common moment of a rudimentary "time" variable that perhaps is not and certainly need not be specified any further. In its moment of sensitivity each detector responds to the appropriate influence: to particle proximity in particle physics, to local field strength in electrodynamics, and to local geometry in geometrodynamics. Whichever the dynamic entity, the measurements of it do not and cannot fully serve dynamic theory unless they span a spacelike hypersurface.

Distill this discussion! Where specify dynamically complete initial value data? On a spacelike 3-geometry, yes; on a null 3-geometry, no. What "points" belong to a topologically acceptable arena for geometrodynamics? Spacelike 3-geometries, yes; null 3-geometries, no. What a remarkable correspondence between dynamics and topology!

Wave Packet in Superspace and Its Propagation

So much for the dynamical entity, space; and so much for the arena in which the dynamics takes place, superspace; now for the dynamics itself. Not one trace of dynamics does one see when he examines the typical probability amplitude function, $\psi = \psi({}^{(3)}\mathcal{G})$, for it is spread all over superspace. No

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

surprise! Already in classical theory the Hamilton–Jacobi function, $S = S({}^{(3)}\mathcal{G})$, is spread out over the manifold. Moreover, this "dynamical phase function" of classical geometrodynamics gives at once in the semiclassical approximation the actual phase of ψ, according to the formula

$$\psi({}^{(3)}\mathcal{G}) = \begin{pmatrix}\text{SLOWLY VARYING}\\ \text{AMPLITUDE FUNCTION}\end{pmatrix} \exp\left(\frac{i}{\hbar}\right) S({}^{(3)}\mathcal{G}) \tag{9}$$

—indication enough that ψ and S are both unlocalized! Dynamics first clearly becomes recognizable when sufficiently many such spread-out probability amplitude functions are superposed to build up a localized wave packet [8, 9]:

$$\psi = c_1\psi_1 + c_2\psi_2 + \cdots \tag{10}$$

Constructive interference occurs where the phases of the several individual waves agree [10]:

$$S_1({}^{(3)}\mathcal{G}) = S_2({}^{(3)}\mathcal{G}) = \cdots \tag{11}$$

The ${}^{(3)}\mathcal{G}$'s compatible with these conditions of constructive interference constitute the classical geometrodynamical history of space (Fig. 3): Marvelously, every 3-geometry that satisfies the conditions of constructive interference (11) can be obtained as a spacelike slice through a certain 4-geometry. More marvelously, this ${}^{(4)}\mathcal{G}$ satisfies the ten field equations of Einstein. In other words, *in addition to the principle of constructive interference one needs only the single equation of Hamilton and Jacobi for the "dynamical phase" $S({}^{(3)}\mathcal{G})$ to obtain all of classical geometrodynamics. The proof of this important point has been announced by Gerlach* [11].

The Hamilton–Jacobi equation itself was first given explicitly in the literature by Peres [12] on the foundation of earlier work by himself and others on the Hamiltonian formulation of geometrodynamics [13]:

$$g^{-1}(g_{ik}g_{jl} - \tfrac{1}{2}g_{ij}g_{kl})\left(\frac{\delta S}{\delta g_{ij}}\right)\left(\frac{\delta S}{\delta g_{kl}}\right) + {}^{(3)}R = 0 \tag{12}$$

Here the g_{ij} are the coefficients in the metric on ${}^{(3)}\mathcal{G}$ and g is the determinant of these metric coefficients. The quantity ${}^{(3)}R$ is the local value of the scalar curvature invariant of the geometry intrinsic to ${}^{(3)}\mathcal{G}$. The Hamilton–Jacobi function depends only upon the 3-geometry, and not upon how that 3-geometry is expressed in terms of metric coefficients in a particular coordinate patch. However, S is treated in (12) as if it depended upon the metric coefficients individually, $S = S(g_{11}, g_{12}, \ldots, g_{33})$. With this understanding, $\delta S/\delta g_{ij}$ denotes the functional derivative of S with respect to alterations in the function $g_{ij}(x, y, z)$. The fact that coordinates in the end have nothing to do

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

with the matter can be stressed by rewriting (12) symbolically in the form

$$\boxed{\left(\frac{\nabla S}{\delta^{(3)}\mathcal{G}}\right)^2 + {}^{(3)}R = 0} \tag{13}$$

This "Einstein–Hamilton–Jacobi equation" contains all of classical geometrodynamics in regions where no "real" sources of mass-energy are present. This one equation carries the entire content of Einstein's ten field equations. Consequently, this equation has been checked and is subject to further check, in the same sense and to the same degree that the predictions of Einstein's field equations have been verified, and are subject to further verification. In the simplest version of particle physics a world line has a simple meaning: It is a stockpile consisting of all those points (x, t) that satisfy the condition of constructive interference. There are ∞^1 of these points on the world line, if we use an obvious though loose way of counting. Dynamics marks out these points and gives preference to them over all the ∞^2 points in the arena of particle dynamics. In geometrodynamics a 4-geometry has a similar significance. It is a stockpile consisting of all those 3-geometries that satisfy the condition of constructive interference. There are ∞^{∞^3} of these 3-geometries, obtainable by making a spacelike slice through the 4-geometry in one or another way:

$$t = t(x, y, z) \tag{14}$$

thus,

(a) ∞^3 points (x, y, z)
(b) ∞^1 choices for t at each of these points

hence

(c) ∞^{∞^3} choices of 3-geometry altogether (15)

Classical geometrodynamics marks out these ${}^{(3)}\mathcal{G}$'s and gives preference to them over the infinitely more numerous totality of ${}^{(3)}\mathcal{G}$'s to be found in the entire arena of geometrodynamics. That arena, superspace, contains ${}^{(3)}\mathcal{G}$'s to the number $(\infty^3)^{\infty^3}$, calculated most simply as follows:

(a) ∞^3 points; and, in a coordinate system that makes the metric diagonal,
(b) 3 diagonal components of the metric specifiable per space point;

and hence

(c) ∞^3 choices of the metric per space point;

and therefore

(d) $(\infty^3)^{\infty^3}$ choices of ${}^{(3)}\mathcal{G}$ altogether (16)

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

From this totality of conceivable 3-geometries one has to pick out and exhibit all the dynamically allowed 3-geometries before he has told in all fullness how space evolves with time. No new lesson! One has long known that time in general relativity is a many-fingered entity. The hypersurface drawn through spacetime to give one ${}^{(3)}\mathcal{G}$ can be pushed forward in time a little here or a little there or a little somewhere else to give one or another or another new ${}^{(3)}\mathcal{G}$. "*Time*" conceived in these terms *means* nothing more or less than *the location of the* ${}^{(3)}\mathcal{G}$ *in the* ${}^{(4)}\mathcal{G}$. In this sense "3-geometry is a carrier of information about time" [14].

"Spacetime," a Concept of Limited Validity

The child's toy can be removed from its box only to reveal another box and—that taken away—another box, and so on, until eventually there are dozens of boxes scattered over the floor. Or conversely the boxes can be put back together, nested one inside the other, to reconstitute the original package. The packaging of ${}^{(3)}\mathcal{G}$'s into a ${}^{(4)}\mathcal{G}$ is much more sophisticated. Nature provides no monotonic ordering of the ${}^{(3)}\mathcal{G}$'s. Two of the dynamically allowed ${}^{(3)}\mathcal{G}$'s taken at random will often cross each other one or more times. When one shakes the ${}^{(4)}\mathcal{G}$ apart, he therefore gets enormously more ${}^{(3)}\mathcal{G}$'s "spread out over the floor" than he might otherwise have imagined. Conversely, when one puts back together all of the ${}^{(3)}\mathcal{G}$'s allowed by the condition of constructive interference, he gets a structure with a rigidity that he might not otherwise have foreseen. This rigidity arises from the infinitely rich interleaving and intercrossing of clear-cut well-defined ${}^{(3)}\mathcal{G}$'s one with another. In summary, (1) the ${}^{(3)}\mathcal{G}$'s allowed by (11) are the basic building blocks; (2) their interconnections give ${}^{(4)}\mathcal{G}$ its existence, its dimensionality [15], and its "magic structure"; and (3) in this structure every ${}^{(3)}\mathcal{G}$ has a rigidly fixed location of its own.

How different from the textbook concept of spacetime! There the geometry of spacetime is conceived as constructed out of elementary objects, or points, known as "events." Here, by contrast, the primary concept is 3-geometry, and the event is secondary: (1) The event lies at the "intersection" of such and such ${}^{(3)}\mathcal{G}$'s. (2) Its timelike relation to some other ${}^{(3)}\mathcal{G}$ is determined by the structure of the ${}^{(4)}\mathcal{G}$, which in turn derives from the intercrossings of all the other ${}^{(3)}\mathcal{G}$'s.

Whether one starts with ${}^{(3)}\mathcal{G}$'s as primary and regards the "event" as a derived concept, or vice versa, might make little difference if one were to remain in the domain of classical geometrodynamics. It makes all the difference when one turns to quantum geometrodynamics.

There is no such thing as a 4-geometry in quantum geometrodynamics, and for a simple reason. No probability amplitude function $\psi({}^{(3)}\mathcal{G})$ can propagate through superspace as an indefinitely sharp wave packet. It

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

spreads (Fig. 3). It has a finite probability amplitude in a domain of superspace of finite measure. This domain encompasses a set of ${}^{(3)}\mathcal{G}$'s far too numerous to accommodate in any one ${}^{(4)}\mathcal{G}$. One can express this situation in various terms. One can say that propagation takes place in superspace, not by following any one classical history of space, not by following any one ${}^{(4)}\mathcal{G}$, but by summation of contributions from an infinite variety of such histories. This extension of Feynman's concept of "sum over histories" has received special attention from Misner [16]. In whatever way one states the matter, however, the facts are clear. The ${}^{(3)}\mathcal{G}$'s that occur with significant probability amplitude do not fit and cannot be fitted into any single ${}^{(4)}\mathcal{G}$. That "magic structure" of classical geometrodynamics simply does not exist. Without that building plan to organize the ${}^{(3)}\mathcal{G}$'s of significance into a definite relationship, one to another, even the "time ordering of events" is a notion devoid of all meaning.

These considerations reveal that the concepts of spacetime and time itself are not primary but secondary ideas in the structure of physical theory. These concepts are valid in the classical approximation. However, they have neither meaning nor application under circumstances when quantum-geometrodynamical effects become important. Then one has to forgo that view of nature in which every event, past, present, or future, occupies its preordained position in a grand catalog called "spacetime." There is no spacetime, there is no time, there is no before, there is no after. The question what happens "next" is without meaning.

The Planck Length and Gravitational Collapse

Under everyday circumstances these unexpected consequences of the quantum principle never come into evidence. The characteristic dimension of quantum geometrodynamics is the Planck length [17], $(\hbar G/c^3)^{1/2} = 1.6 \times 10^{-33}$ cm. By comparison the normally relevant scale of any geometry of interest is stupendous. Negligible on the scale of the geometry is the quantum-mechanical spread of the wave packet in superspace. Consequently the dynamical evolution of the geometry can be treated in the context of classical geometrodynamics. Thus the geometries that occur with significant probability amplitude can be idealized to a good approximation as if confined to a region in superspace of zero thickness. The ${}^{(3)}\mathcal{G}$'s of this limited set are sufficiently small in number to fit together into a single ${}^{(4)}\mathcal{G}$. They are sufficiently large in number to reproduce every conceivable spacelike slice of that ${}^{(4)}\mathcal{G}$. In this approximation it makes good sense to speak of "the classical geometrodynamical history of space."

If the dynamics of geometry thus normally lends itself to classical analysis, there are two contexts where it does not. One is the final stage of

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

gravitational collapse. The other is analysis of the microscopic quantum-mechanical fluctuations in the geometry of space and their consequences for physics generally.

Of all the applications of quantum geometrodynamics, none would seem more immediate than gravitational collapse [18]. Here according to classical general relativity the dimensions of the collapsing system in a finite proper time are driven down to indefinitely small values. The phenomenon is not limited to the space occupied by matter. It occurs also in the space surrounding the matter. In a finite proper time the calculated curvature rises to infinity. At this point classical theory becomes incapable of further prediction. In actuality, classical considerations go wrong before this point. A prediction that is infinity is not a prediction. The wave packet in superspace does not and cannot follow the classical history when the geometry becomes smaller in scale than the quantum-mechanical spread of the wave packet. Not a new phenomenon! Throughout physics one sees examples of a wave impinging upon a region of interaction of dimensions small compared to a wavelength. The outcome is scattering or diffraction. One speaks of a probability for this, that, or the other outcome of the interaction. The photon or phonon or other entity entering from one direction emerges in another. The concept of a deterministic world line may serve adequately during the phase of approach to the zone of interaction, and during the phase of regression. It is completely out of place during the phase of scattering. So here. The concept of a deterministic history of geometry, a well-defined ${}^{(4)}\mathcal{G}$, makes sense in the early phase of gravitational collapse, but has absolutely no application in the decisive phase. There "space-time" is nonexistent, "events" and the "time ordering of events" are without meaning, and the question "what happens after the final phase of gravitational collapse" is a mistaken way of speaking.

The correct way of speaking deals with the propagation of the probability amplitude $\psi({}^{(3)}\mathcal{G})$ in superspace. The semiclassical treatment of the propagation (Eqs. (9), (13)) is appropriate in most of the domain of superspace of interest for gravitational collapse. Not so in the decisive region. There, as in elementary problems of scattering, the mathematical analysis has to go straight back to the full and accurate wave equation for its foundation. What is the appropriate question to ask of the mathematics? One has learned how to formulate the right question, in the case of scattering, through long experience with the physics. However, if one had not had that physical experience, he would have learned out of the mathematical formalism itself to speak about incoming plane waves and outgoing spherical waves, and scattering amplitudes, and about all the other relevant concepts. Similarly in quantum geometrodynamics, where one has so much less experience. The mathematical formalism itself must serve as the final arbiter on how to pose the central question about gravitational collapse as well as how to answer it!

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Quantum Fluctuations in Geometry of Space

"Quantum fluctuations in the geometry of space" is the other pressing field of application of quantum geometrodynamics. What a strange combination of words! Fluctuations are well known. The term "quantum fluctuations" carries a deeper meaning. It stands for a movement that can never be frozen out, however low the temperature. Such fluctuations are universal. In the hydrogen molecule both the separation of the two atoms and their relative momentum continually flucuate. Fixity of both would violate the uncertainty principle. In the frozen vaccuum of quantum electrodynamics the electric and magnetic fields both fluctuate. Were both of these dynamically conjugate field variables to vanish, the uncertainty principle would likewise fail. The same is true of quantum geometrodynamics. There the conjugate variables are the "intrinsic curvature" of three-dimensional space and the "extrinsic curvature," telling how this space is bent relative to the geometry of any 4-dimensional space-time that might envelop it. For both dynamic quantities to be stilled would equally contradict Heisenberg's uncertainty relation. Thus all space at the quantum scale of distances is the seat of the liveliest geometrodynamics, as it is also everywhere the scene of the most violent small-scale fluctuations in the electromagnetic field.

No prediction of quantum electrodynamics has been more impressively verified in the whole post-World War II era than these vacuum fluctuations in the electric field. Their perturbing influence on the motion of the electron (Fig. 4) accounts for the major component in the Lamb shift in the energy levels of the hydrogen atom [19].

In putting numbers to the fluctuations in geometry one is guided by the example of electromagnetism and, at a still earlier stage, by the example of the harmonic oscillator. For the oscillator the expression for the energy is

$$E = \left\langle \begin{pmatrix}\text{kinetic}\\ \text{energy}\end{pmatrix} + \begin{pmatrix}\text{potential}\\ \text{energy}\end{pmatrix} \right\rangle = \frac{p^2}{2m} + \tfrac{1}{2}m\omega^2 x^2$$

$$= \int \psi^*(x)\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \tfrac{1}{2}m\omega^2 x^2 \right]\psi(x)\,dx \tag{17}$$

For a probability amplitude function $\psi(x)$ of Gaussian form and range a,

$$\psi(x) = \pi^{-1/4}a^{-1/2}\exp\left(-\frac{x^2}{2a^2}\right) \tag{18}$$

the expectation value of the energy is

$$E = \frac{\hbar^2}{4ma^2} + \tfrac{1}{4}m\omega^2 a^2 \tag{19}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

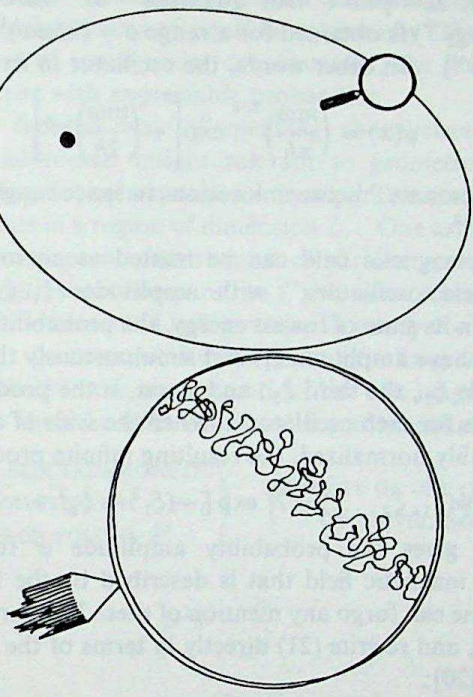


FIGURE 4. Symbolic representation of motion of electron in hydrogen atom as affected by fluctuations in electric field in vacuum ("vacuum" or "ground state" or "zero-point" fluctuations). The electric field associated with the fluctuation, $E_x(t) = \int E_x(\omega) e^{-i\omega t}\, d\omega$, brings about in the most elementary approximation the displacement $\Delta x = \int (e/m\omega^2) E_x(\omega) e^{-i\omega t}\, d\omega$. The average vanishes but the root mean square $\langle (\Delta x)^2 \rangle$ does not. In consequence the electron feels an effective atomic potential altered from the expected value $V(x, y, z)$ by the amount

$$\Delta V(x, y, z) = \tfrac{1}{2} \langle (\Delta x)^2 \rangle \nabla^2 V(x, y, z)$$

The average of this perturbation over the unperturbed motion accounts for the major part of the observed Lamb–Rutherford shift $\Delta E = \langle \Delta V(x, y, z) \rangle$ in the energy level. Conversely, the observation of the expected shift makes the reality of the vacuum fluctuations inescapably evident.

If quantum effects were absent, the first term would disappear. The minimum energy would be obtained by putting the oscillator at rest at the origin. However, in the real world of quantum physics such a sharp localization in position ($a = 0$) would make the effective wavelength zero and the momentum and kinetic energy arbitrarily large (divergent first term in Eq. (19)). The

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

minimum energy, $E = \frac{1}{2}\hbar\omega$ ("half quantum" or "zero-point energy" or "fluctuation energy") is obtained for a range $a = (\hbar/m\omega)^{1/2}$ ("range of zero-point oscillations"). In other words, the oscillator in its ground state

$$\psi(x) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \exp\left[-\left(\frac{m\omega}{2\hbar}\right)x^2\right] \tag{20}$$

can be said to "resonate" between locations in space ranging over a region of extent $\sim(\hbar/m\omega)^{1/2}$.

The electromagnetic field can be treated as an infinite collection of independent "field oscillators," with amplitudes $\xi_1, \xi_2, \ldots$. When the Maxwell field is in its state of lowest energy, the probability amplitude for the first oscillator to have amplitude ξ_1, and simultaneously the second oscillator to have amplitude ξ_2, the third ξ_3, and so on, is the product of functions of the form (20), one for each oscillator. When the scale of amplitudes for each oscillator is suitably normalized, the resulting infinite product takes the form

$$\psi(\xi_1, \xi_2, \ldots) = N \exp\left[-(\xi_1^2 + \xi_2^2 + \cdots)\right] \tag{21}$$

This expression gives the probability amplitude ψ for a configuration $\mathbf{B}(x, y, z)$ of the magnetic field that is described by the Fourier coefficients $\xi_1, \xi_2, \ldots$. One can forgo any mention of these Fourier coefficients if he so desires, however, and rewrite (21) directly in terms of the magnetic field configuration itself [20]:

$$\psi(\mathbf{B}(x, y, z)) = \mathcal{N} \exp\left(-\iint \frac{\mathbf{B}(x_1)\cdot\mathbf{B}(x_2)}{16\pi^3\hbar c r_{12}^2}\right) d^3x_1\, d^3x_2 \tag{22}$$

No longer does one speak of "the" magnetic field; he talks instead of the probability of this, that, or the other configuration of the magnetic field and this even under circumstances, as here, where the electromagnetic field is in its ground state.

It is reasonable enough under these circumstances that the configuration of greatest probability is $\mathbf{B}(x, y, z) = 0$. Consider for comparison a configuration where the magnetic field is again everywhere zero except in a region of dimension L. There let the field, subject as always to the condition div $\mathbf{B} = 0$, be of the order of magnitude ΔB. The probability amplitude for this configuration will be reduced relative to the nil configuration by a factor $\exp(-I)$. Here the quantity I in the exponent is of the order $(\Delta B)^2 L^4/\hbar c$. Configurations for which I is large compared to 1 occur with negligible probability. Configurations for which I is small compared to 1 occur with practically the same probability as the nil configuration. In this sense, one can say that the fluctuations in the magnetic field in a region of extension L are of the order of magnitude

$$\Delta B \sim \frac{(\hbar c)^{1/2}}{L^2} \tag{23}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

In other words, the field "resonates" between one configuration and another with the range of configurations of significance given by (23). Moreover, the smaller is the region of space under consideration, the larger are the field magnitudes that occur with appreciable probability.

Still another familiar way of speaking about electromagnetic field fluctuations gives additional insight relevant to geometrodynamics. One considers a measuring device responsive in comparable measure to the magnetic field at all points in a region of dimension L. One asks for the effect on this device of electromagnetic disturbances of various wavelengths. A disturbance of wavelength short compared to L will cause forces to act one way in some parts of the detector and will give rise to nearly compensating forces in other parts of it. In contrast, a disturbance of a long wavelength λ produces forces everywhere in the same direction, but of a magnitude too low to have much effect. Thus the field, estimated from the equation

$$\begin{pmatrix}\text{ENERGY OF ELECTROMAGNETIC}\\ \text{WAVE OF WAVELENGTH } \lambda \text{ IN A}\\ \text{DOMAIN OF VOLUME } \lambda^3\end{pmatrix} \sim \begin{pmatrix}\text{ENERGY OF ONE QUANTUM}\\ \text{OF WAVELENGTH } \lambda\end{pmatrix}$$

or

$$B^2\lambda^3 \sim \frac{\hbar c}{\lambda}$$

or

$$B \sim \frac{(\hbar c)^{1/2}}{\lambda^2} \tag{24}$$

is very small if λ is large compared to the domain size L. The biggest effect is caused by a disturbance of wavelength λ comparable to L itself. This line of reasoning leads directly from (24) to the standard fluctuation formula (23).

Fluctuations Superposed on Classical Background

Nothing says that the electromagnetic field has to be in its ground state. It can be excited by a distant wireless antenna so that locally it is oscillating up and down at 10^6 cycles/sec in the range $B = \pm 3.3 \times 10^{-8}$ gauss, for example (accompanying electric fields 1 millivolt/meter). By comparison with this deterministic classical history of the local magnetic field,

$$B_z = 3.3 \times 10^{-8} \text{ gauss} \cos 10^6 t \tag{25}$$

the quantum-mechanical fluctuation field of the same frequency (Eq. (24)) is quite negligible:

$$B \sim \frac{6 \times 10^{-9}(\text{erg cm})^{1/2}}{(3 \times 10^4 \text{ cm})^2}$$

$$\sim 10^{-17} \text{ gauss} \tag{26}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Even the total effect of all the independent quantum fluctuations in the magnetic field is small as sensed by a detector of dimension $L \sim 1$ cm; thus, from Eq. (23),

$$\Delta B \sim \frac{6 \times 10^{-9}(\text{erg cm})^{1/2}}{(1 \text{ cm})^2}$$

$$\sim 6 \times 10^{-9} \text{ gauss} \tag{27}$$

However, when attention is fixed on the field in a still smaller domain, say $L \sim 0.1$ cm or less, then the quantum fluctuations in the magnetic field dominate over the deterministic classical field of (25):

$$\Delta B > 6 \times 10^{-7} \text{ gauss} \tag{28}$$

So much for the coexistence of classical fields and quantum fluctuations in electrodynamics!

Similar considerations apply in geometrodynamics [21]. Quantum fluctuations in the geometry are superposed on and coexist with the large-scale slowly varying curvature predicted by classical deterministic general relativity. Thus, in a region of dimension L, where in a local Lorentz frame the normal values of the metric coefficients will be $-1, 1, 1, 1$, there will occur fluctuations in these coefficients of the order

$$\Delta g \sim \frac{L^*}{L} \tag{29}$$

fluctuations in the first derivatives of the g_{ik}'s of the order

$$\Delta \Gamma \sim \frac{\Delta g}{L} \sim \frac{L^*}{L^2} \tag{30}$$

and fluctuations in the curvature of space of the order

$$\Delta R \sim \frac{\Delta g}{L^2} \sim \frac{L^*}{L^3} \tag{31}$$

Here

$$L^* = \left(\frac{\hbar G}{c^3}\right)^{1/2} = 1.6 \times 10^{-33} \text{ cm} \tag{32}$$

is the so-called [22] Planck length. It is appropriate to look at orders of magnitude. The curvature of space within and near the earth, according to

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

classical Einstein theory, is of the order

$$R \sim \left(\frac{G}{c^2}\right)\rho \sim (0.7 \times 10^{-28}\ \text{cm/g})(5\ \text{g/cm}^3)$$

$$\sim 4 \times 10^{-28}\ \text{cm}^{-2} \tag{33}$$

This quantity has a very direct physical significance. It measures the "tide-producing component of the gravitational field" as sensed, for example, in a freely falling elevator or in a space ship in free orbit around the earth [23]. By comparison the quantum fluctuations in the curvature of space are only

$$\Delta R \sim 10^{-33}\ \text{cm}^{-2} \tag{34}$$

even in a domain of observation as small as 1 cm in extent. Thus the quantum fluctuations in the geometry of space are completely negligible under everyday circumstances.

Even in atomic and nuclear physics the fluctuations in the metric,

$$\Delta g \sim \frac{10^{-33}\ \text{cm}}{10^{-8}\ \text{cm}} \sim 10^{-25}$$

and

$$\Delta g \sim \frac{10^{-33}\ \text{cm}}{10^{-13}\ \text{cm}} \sim 10^{-20} \tag{35}$$

are so small that it is completely in order to idealize the physics as taking place in a flat Lorentzian spacetime manifold.

The quantum fluctuations in the geometry are nevertheless inescapable, if we are to believe the quantum principle and Einstein's theory. They coexist with the geometrodynamical development predicted by classical general relativity. The fluctuations widen the narrow swathe cut through superspace by the classical history of the geometry. In other words, the geometry is not deterministic, even though it looks so at the everyday scale of observation. Instead, at a submicroscopic scale it "resonates" between one configuration and another and another. This terminology means no more and no less than the following: (1) Each configuration ${}^{(3)}\mathcal{G}$ has its own probability amplitude $\psi = \psi({}^{(3)}\mathcal{G})$. (2) These probability amplitudes have comparable magnitudes for a whole range of 3-geometries included within the limits (29) on either side of the classical swathe through superspace. (3) This range of 3-geometries is far too variegated on the submicroscopic scale to fit into any one 4-geometry, or any one classical geometrodynamical history. (4) Only when one overlooks these small-scale fluctuations ($\sim 10^{-33}$ cm) and examines the larger-scale features of the 3-geometries do they appear to fit into a single space-time manifold, such as comports with the classical field equations.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Extrapolate Geometrodynamics to the Planck Scale of Distances?

Is it not preposterous to apply existing theory in a realm of dimensions smaller than nuclear sizes by twenty powers of ten? What a fantastic extrapolation! Yet it is the tradition of theoretical physics to adopt what might be called a "strong bargaining posture." It is not the custom to give up any long-established principle without pushing it to the limit and finding out where, if anywhere, it goes wrong. A direct contradiction between a prediction and an observation, or between two points of principle, is ordinarily necessary if one is to have any solid ground for change, or even any indication where to make a change! The physicist does not have the habit of giving up something unless he gets something better in return.

To pursue systematically the consequences of quantum geometrodynamics is recommended not merely by the absence of any contradiction and by the absence of any more comprehensive theory. It is also made attractive by an example out of the past. Who in the 1850's, measuring the attraction between electric charges and testing the Coulomb law at distances from meters to millimeters, could have predicted that it would be proved valid in 1911 to 10^{-12} cm, in 1933 to 10^{-13} cm, and in 1963 to 10^{-14} cm? The fantastic extrapolatory power of basic physical theory always seems a miracle [24]!

Electrodynamics contains no natural length. Neither do general relativity (G, c) or the quantum principle ($\hbar$) individually. The union of geometrodynamics and the quantum does: $L^* = (\hbar G/c^3)^{1/2}$. The importance of what is essentially this length was stressed by Planck as early as 1899 [25]. He had taken up the study of blackbody radiation not least because it is universal: independent of the shape and size of the container, independent of the properties of its walls, and independent of the complexities of atomic, molecular, and solid state physics. In keeping with this search for the universal, he asked for standards of length, mass, and time that are independent of such special circumstances as the size of the planet we happen to inhabit, its period of rotation, and the density of the fluid that covers it. In excluding reference to special substances he found it natural also to exclude reference to special particles: both the electron, with its then known mass and charge, and all heavier entities—objects that even today pose unsolved structural problems. Excluding all else, he was left with the speed of light, the Newtonian constant of gravitation, and the constant newly discovered from the analysis of blackbody radiation as the quantities he was willing to accept as truly fundamental. Out of these three quantities there is but one way to construct a length. Planck's length, introduced to science before either special or general relativity, first acquired an understandable role in the context of quantum geometrodynamics as measure of the fluctuations in the geometry of space.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Accept seriously a length as small as 10^{-33} cm? Try to assess in any detail at all the physics that goes on at a scale of distances shorter by twenty powers of ten than the 10^{-13} cm of elementary particle physics? What could be more preposterous? Only three numbers are still more preposterous than 10^{20}: the factor of 10^{40} that distinguishes electric forces from gravitational forces; the 10^{40} from elementary particle dimensions to the estimated radius of the universe at the phase of maximum expansion; and the 10^{80} that furnishes an order of magnitude estimate as good as any that one knows for the number of particles in the universe. Eddington [26], Dirac [27], Jordan [28], Dicke [29], and Hayakawa [30] argue that it is unreasonable to think of such enormous numbers as having independent roles in physics. The correspondence between these numbers cannot be purely accidental, they stress [31]; there could hardly be a "regularity of the large numbers" if there were not a deep connection between cosmology, general relativity, and elementary particle physics. But where to begin in looking for this connection? Begin with asking why there are so and so many particles in the universe? Hardly. Physics can elucidate laws of motion, but it has proved powerless to explain initial conditions. Begin with asking why the universe has such and such dimensions? Again a matter of initial conditions, outside of the present scope of physics. Begin with trying to explain the charge structure of elementary particles, or the characteristic dimension of 10^{-13} cm? More hopeful, perhaps, but still beyond present power. No, of all the quantities coupled by the large numbers, one alone has a clear status within existing theory: the Planck length. Where else than here can one begin?

It could seem plausible to stop considering physics a little below 10^{-13} cm because accelerator budgets stop a little above $100 million a year. What good is it, one sometimes asks, to analyze what goes on if one has no way to observe it? Happily, experience has taught views less biased by mankind's temporary limitations. Wait until one had mastered the work hardening of metals before one took up the microscope and saw dislocations? Wait until one had explained dislocations before one started the study of atoms? Not so. The route to understanding did not go down the ladder, 1 cm $\rightarrow$ 10^{-4} cm $\rightarrow$ 10^{-8} cm, but up: 10^{-8} cm $\rightarrow$ 10^{-4} cm $\rightarrow$ 1 cm. One had to understand something about atoms before he could explain dislocations, and something about dislocations before he could uncover the rationale of work hardening. Is it possible that one similarly must have some perspective on what happens at 10^{-33} cm before one can find the rationale of particles and 10^{-13} cm [32]? Right or wrong [33], quantum geometrodynamics alone has any suggestions to offer on this point. What then does it say?

Every new perspective offered by this long established theory radiates out from the central prediction: *geometry fluctuates violently at small distances.* This concept opens new views on the nature of electric charge, on the nature of the vacuum, and on the nature of particles.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ELECTRICITY AS LINES OF FORCE TRAPPED IN THE TOPOLOGY OF SPACE

To arrive at a new vision of electricity it is enough to question an old view of topology: "Space is Euclidean in character at small distances." The view is reasonable enough for everyday purposes. Equally reasonable is the conception of the surface of the ocean as endowed with Euclidean topology—reasonable to one flying miles above it. To one in a small boat the opposite impression is inescapable. He sees the breaking waves and the foam. He knows that the surface is multiply connected at the scale of millimeters and centimeters. If the ocean is violent, the geometry of space on the Planck scale of distances is even more violent. Nowhere is there any region of calm. Moreover, if the equations of hydrodynamics are nonlinear, so are the equations of geometrodynamics. What a contrast to the linearity of electrodynamics! There the predicted fluctuations in potential

$$\Delta A \sim \frac{(\hbar c)^{1/2}}{L} \tag{36}$$

and in field

$$\Delta F \sim \frac{(\hbar c)^{1/2}}{L^2} \tag{37}$$

preserve always the same character, regardless of the smallness of the distances L to which one goes in his probing. There is no natural magnitude to mark off large fluctuations as different in nature from small ones. The contrary is the case for fluctuations in the metric, governed by the formula

$$\Delta g \sim \frac{L^*}{L} \tag{38}$$

Values of Δg comparable to unity and larger indicate changes in geometry so drastic that the word "curved space" is hardly adequate to describe them. "Changes in topology" seems a more reasonable description.

It is not so natural in mathematics as in physics to consider a transformation that alters one topology to another. An oscillating drop of water undergoes fission. The topology changes. A point marks the place of separation of the two masses of liquid. That point lacks the full neighborhood of points that characterizes a normal point. Such a critical point is ruled out from any proper manifold by the very definition of the term "manifold" in mathematics. Before the division, the surface of the drop constituted a manifold. After the division, it is again a manifold, consisting of two disparate pieces. At the instant of division it is not a manifold. But little

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

attention does the drop pay to this distinction. It divides, despite all definitions. No more reason does one see in the definition of "manifold" against *space* changing *its* topology.

No principle is at hand that would give one topology perpetual preference over all others. On the contrary, the field equations of relativity are purely local in character. They make no statements at all about global topology, as Einstein himself emphasized more than once. Moreover, the whole character of physics speaks for the theme that "everything that can happen will happen." An alpha particle penetrates through a region classically forbidden to it; the side group on a chain molecule undergoes "hindered rotation"; and the umbrella structure of an ammonia molecule turns inside out despite the apparent contradiction to the law of conservation of energy. It is difficult to resist the conclusion that likewise the topology of space can change and does change.

If these general considerations are relevant, and if fluctuations alter the topology of space as well as its curvature [34], the consequences are decisive for the nature of the physics that goes on at small distances, and even for the nature of superspace itself. Superspace has to be broadened from the totality of positive definite 3-geometries built on one topology to the totality of positive definite 3-geometries built on the totality of all topologies. It has new implications to say that the probability amplitude $\psi({}^{(3)}\mathcal{G})$ is appreciable for a swathe of points in superspace, with a finite spread about the deterministic history of classical geometrodynamics (Fig. 5): Geometry in the small fluctuates not only from one microscopic pattern of curvature to another, but much more, from one microscopic topology to another. Moreover, those structures that are everywhere full of submicroscopic "handles" or "wormholes" are overwhelmingly more numerous than 3-geometries of simpler topology. In other words, space "resonates" between one foamlike structure and another [35]. The space of quantum geometrodynamics can be compared to a carpet of foam spread over a slowly undulating landscape. The undulations symbolize deterministic classical geometrodynamics. The continual microscopic changes in the carpet of foam as new bubbles appear and old ones disappear symbolize the quantum fluctuations in the geometry. The fluctuations change the microscopic connectivity of space itself. No longer is one entitled to take it for granted that space is Euclidean in the small.

Nowhere in physics does the structure of geometry in the small play a larger part than in electricity. Electric lines of force converge onto a region of space and none come out of it. Something strange must go on in that region. Either Maxwell's equations break down, or the region is filled with a special substance, an electric jelly, a magic fluid beyond further explanation. From the one picture or the other there has never been an escape. The reason is simple. The region is tacitly assumed to have Euclidean topology. Give up this assumption [36], but hold to Maxwell's field equations for empty

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

space. Then the conclusion changes. The region in question must contain the mouth of at least one "handle" or "wormhole" (Fig. 6). Electric lines of force converge upon this mouth only to emerge from the other mouth, located somewhere else in space. One comes in this way to a new picture of electricity: *A classical geometrodynamical electric charge is a set of lines of force trapped in the topology of space.*

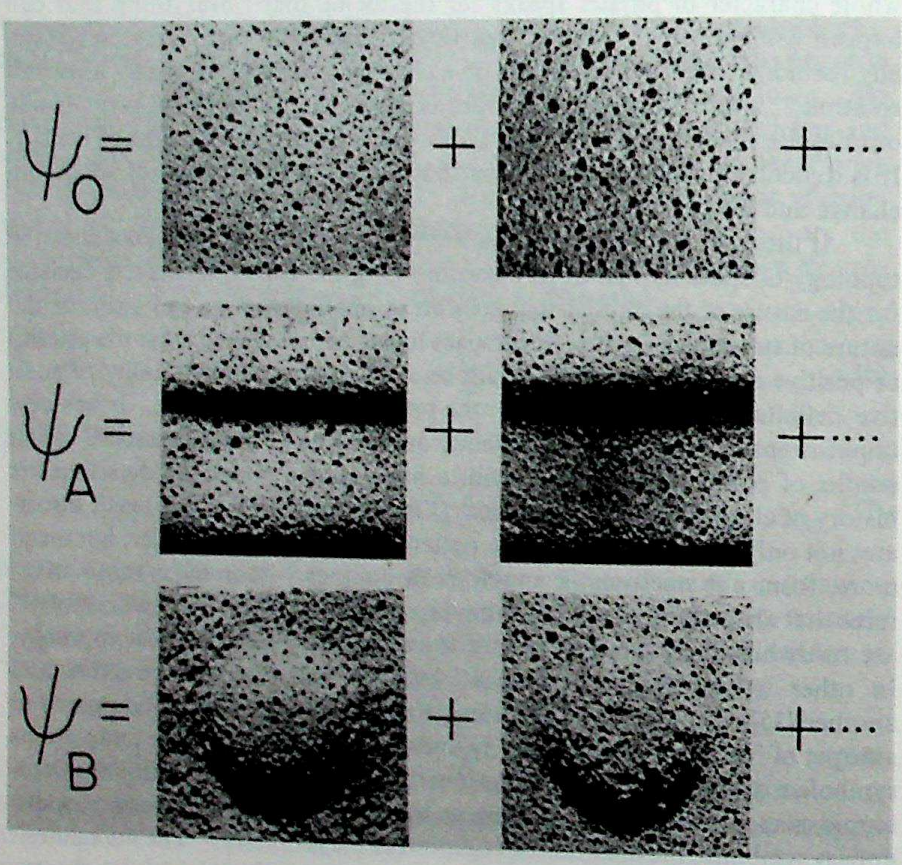


FIGURE 5. Symbolic representations of three alternative probability functions $\psi({}^{(3)}\mathcal{G})$. Above, ψ_0, normal fluctuations alone; middle, ψ_A, macroscopic classical gravitational wave plus superposed fluctuations; below, ψ_B, localized excitation plus superposed fluctuations. The complex number ψ gives the probability amplitude for the occurrence of the 3-geometry ${}^{(3)}\mathcal{G}$. Those 3-geometries that contribute most to the totalized probability are highly multiply connected at the scale of the Planck length ("foamlike structure of space in the small").

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Someone can be imagined who first studies topology, next takes to heart Maxwell's equations for empty space, and then for the first time sees an electric charge. He takes it as experimental evidence that space must be multiply connected in the small (Table 2). Nothing prevents us from adopting the same point of view. On this view, *the occurrence of electric charges in nature is the single most impressive piece of evidence available today for the reality of the fluctuations that quantum theory predicts in the geometry of space, and suggests in the topology of space, at the Planck scale of distances.*

The "wormholes" predicted by quantum geometrodynamics are a property of all space, are submicroscopic, and they and the fluxes through

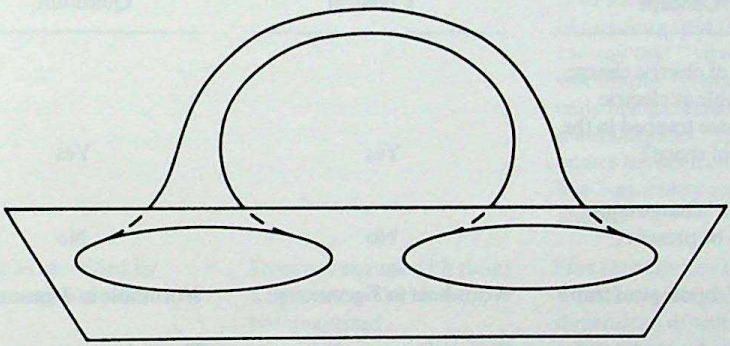


FIGURE 6. Classical geometrodynamical concept of charge as "electric lines of force trapped in the topology of space." The "wormhole" connects two regions in one otherwise nearly Eucildean space, not two different Euclidean spaces. The distance between the two mouths of the wormhole (1) via the nearly Euclidean space and (2) via the route through the wormhole are permitted by Einstein's field equations to be quite different, even in order of magnitude, contrary to what one might assume from the figure. There the geometry (one dimension suppressed!) is depicted for simplicity as if embedded in flat Euclidean 3-space. The third dimension (distance "off" of the geometry) is to be considered as unattainable as the stratosphere is unattainable to an ant crawling on the surface of the earth. An observer endowed with an instrument of inadequate resolving power sees one wormhole mouth as a positive charge, the other as a negative charge. To be contrasted with this classical picture of single indentifiable wormholes in the quantum-mechanical picture (Fig. 4) of a submicroscopic foamlike wormhole structure constantly fluctuating throughout all space.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

them arise spontaneously, through quantum fluctuations. Nothing prevents one from considering also a single wormhole, of macroscopic dimensions, created *ab initio*, with a prescribed flux threading through it, and evolving deterministically in time in accordance with the classical field equations. However, this classical electric charge has not the slightest direct connection with the charges of the real world of quantum physics and requires no consideration here.

TABLE 2 *The Concepts of Electric Charge in Quantum and Classical Geometrodynamics, Compared and Contrasted*

Concept	Classical	Quantum
This type of electric charge interpretable as electric lines of force trapped in the topology of space?	Yes	Yes
Any "real" charge considered to be present?	No	No
Nature of topological trap?	Wormhole in 3-geometry	Wormhole in 3-geometry
Where are wormholes located?	Connecting opposite charges	Throughout all space
Dimension of wormhole	Enormous compared to 10^{-33} cm	Comparable to 10^{-33} cm
Status of wormhole?	Classical deterministic development with time; macroscopic size; this size determined by initial conditions; no observational evidence that any such macroscopic wormhole ever occurred, although by definition it is one of the solutions of Einstein's field equations	"Fluctuation"; consequence of fact geometry is not deterministic; that is, a consequence of the nonzero probability that exists for quite diverse 3-geometries; 3-geometries with wormholes almost everywhere (scale of 10^{-33} cm) are the most numerous ("foamlike 3-geometry"); wormholes do not have to be initiated; they occur naturally ("zero-point disturbance of the vacuum") and cannot be avoided

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

TABLE 2—(*Continued*)

Concept	Classical	Quantum
Wormhole pinches off?	Yes. Throat of radius a undergoes gravitational collapse in time of order a/c (10^{-10} sec for $a = 3$ cm)	Question strictly speaking is undefined; no meaning to deterministic small-scale geometrodynamics in context of quantum geometrodynamics; but in the loose way of speaking that is so useful in certain parts of quantum field theory one can say that "virtual wormholes" are continually being created and annihilated as "space resonates between one foam-like 3-geometry and another"
Charge as identified by flux through wormhole?	Does not change with time; a constant of the motion; not quantized	Flux through any one wormhole, like shape and dimensions of wormhole, subject to quantum-mechanical fluctuations; order of magnitude of this "fluctuation charge," $(\hbar c)^{1/2} \sim 12e$; not "constant" and not "quantized"
Relation of charge on an elementary particle to this kind of geometrodynamical charge?	Not the slightest direct connection with classical geometrodynamical electric charge	Not the slightest direct connection with the charge on an elementary "fluctuation wormhole." Particle viewed as a quantum state of collective excitation of the entire geometrical continuum ("geometrodynamical exciton"). Charge associated with this exciton to be calculated when one understands how to do the calculation!

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

THE ENERGY OF THE VACUUM

If one insight out of fluctuations in geometry has to do with the nature of electricity, a second has to do with the energy of the vacuum. Already in quantum electrodynamics one has long known from both observation and theory that when one examines a region of the vacuum of dimension L (1) the fluctuation energy is found to be of the order $\hbar c/L$, and (2) the effective density of energy of the fluctuation is found to be of the order $\hbar c/L^4$. Not known out of electrodynamics is any natural lower limit for the L values that come into consideration. In quantum geometrodynamics formulas of the same type hold, but there is now a natural cutoff: the Planck length. It is unreasonable to apply linear theory when L is of the order of $L^* = (\hbar G/c^3)^{1/2} = 1.6 \times 10^{-33}$ or less. In other words, there is a certain sense in which one can say: (1) Elementary fluctuations measure in energy up to $\hbar c/L^*$. The magnitude of this characteristic mass-energy is $\sim 10^{-5}$ g or $\sim 10^{28}$ eV; that is, about twenty powers of ten greater than the mass of an elementary particle and nine powers of ten greater than the energy of the most energetic cosmic ray that has ever been found. (2) These fluctuations take place throughout all space. (3) The density of the electromagnetic energy associated with these fluctuations [37] is of the characteristic order of magnitude $\hbar c/L^4 = c^5/\hbar G^2 \sim 10^{95}$ g/cm^3, stupendous in comparison with the 10^{14} g/cm^3 of nuclear matter. (4) The effective density of the "gravitational" or geometrodynamical wave energy [38] associated with these fluctuations is of the same order of magnitude.

Every observation shows that the net density of energy in space is negligible by comparison with these huge figures. The enormous positive energy must be compensated in some way. How the compensation comes about was a major concern of Niels Bohr over the years. A new approach to the problem shows up when one considers the fluctuations at the Planck scale of distances. Two such fluctuations, each of mass-energy $\sim(\hbar c/G)^{1/2} \sim 10^{-5}$ g, interacting gravitationally at the distance L^*, have a coupling energy that is negative,

$$E_{\text{grav}} = -\frac{Gm_1m_2}{r_{12}} \sim -\frac{G\left[\left(\frac{\hbar c}{G}\right)^{1/2}\right]^2}{L^*} \sim -\frac{\hbar c}{L^*} \tag{39}$$

and of the order -10^{-5} g. This coupling between neighboring "fluctuons" thus has such a sign and such a magnitude as to be appropriate for compensating the energies of the individual fluctuons. It would seem surprising if this mechanism did not play a dominant part in bringing about compensation of vacuum energies.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Despite all the mysteries that enshroud the compensation process, one conclusion stands out clear: *Individually the components of the vacuum energy are enormous, and collectively they compensate.*

A PARTICLE AS A GEOMETRODYNAMICAL EXCITON

Quantum fluctuations in geometry offer not only a new picture of electricity and a new view of the violence of the vacuum, but also a third vista: the concept of a particle as a quantum state of excitation of the geometry of space.

A bit of nuclear matter with its density of $\sim 10^{14}$ g/cm^3 is completely unimportant compared to the calculated density of $\sim 10^{95}$ g/cm^3 of the fluctuation energy of the vacuum. A particle means less to the physics of the vacuum than a cloud (10^{-6} g/cm^3) means to the physics of the sky (10^{-3} g/cm^3). No single fact points more powerfully than this to the conclusion that a "particle" is not the right starting point for the description of nature.

From the standpoint of geometrodynamics the primordial entity is not one particle, nor an intercoupled family of particle fields, but the geometry of empty space itself. On this view a particle is not itself a 10^{-33}-cm fluctuation in the geometry; instead, it is a fantastically weak alteration in the pattern of these fluctuations, extending over a zone containing very many such 10^{-33} regions. In brief, a particle is a quantum state of excitation of the geometry; it is a *geometrodynamical exciton.* In mathematical terms, the vacuum is described by one probability amplitude $\psi_0 = \psi_0({}^{(3)}\mathcal{G})$; and states where one or more particles are present are described by other functionals $\psi = \psi({}^{(3)}\mathcal{G})$.

On this interpretation elementary particle physics ranks as a new and beautiful kind of chemistry. First came the chemistry of atoms and molecules, marvelous in its complexities and also in its regularities, all built on one single simple dynamical entity, the electron. Then came "nuclear chemistry," with nuclear shapes and energies and reaction rates all going back for their explanation to the dynamics of another elementary dynamical entity, the nucleon. Today we deal in effect with a chemistry of the elementary particles themselves. Wonderful advances in the subject classify the particles into families, tie their masses together into mathematical regularities, and systematize their rates of transformation—without, however, revealing the identity of the elementary dynamical entity beneath it all. That entity, on the present view, is geometry itself.

When in the first half of the nineteenth century Berzelius proposed that chemical forces are a manifestation of electrical forces [39], he excited investigations by many workers that eventually discredited his hypothesis. The homopolar bond: how could the observed affinity be reconciled with the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

known repulsion between like electric charges? Homopolar forces, ionic forces, van der Waals forces, valence forces: how could all this variety of magnitudes and particularities possibly be compatible with electrical forces, pure and simple? The tide against the electrical interpretation of chemical forces only turned with the discovery of the electron by J. J. Thomson in 1897. The dynamical entity once identified, the unraveling of the mystery eventually had to follow. Still it was not easy for the imagination to grasp what organizing power the quantum principle possesses. In encounters in the mid 1920's more than one physicist told his colleague from the laboratory across the way, "Your chemistry is now passé. All that jumble can now be explained in terms of electrons and quantum numbers." In more than one case the then justified reply came back, "What makes you think your circular and elliptic orbits have anything to do with chemistry? Have you ever heard of the valence angles of ammonia or the tetrahedral bonds of carbon? Don't ever forget that electrical forces are electrical forces and chemical forces are chemical forces." The Coulomb law had to be supplemented by the concept of probability amplitude before Heitler and London could explain valence forces. Today no one doubts that the Schroedinger equation accounts in principle for all of chemistry. Yet no surer way could be found to stop the advance of chemistry than to require everyone to calculate the wave function of his new compound before making it. Not the contemplation of 600-dimensional configuration space, but the analysis of the regularities between molecule and molecule, proves to be the fruitful way to make progress. It can hardly be otherwise when the energy of binding is the very small difference between the very much larger total energies of the associated and dissociated states.

In "elementary particle chemistry" the art of analyzing regularities is already highly advanced [40], thanks not least to applications of group theory even more far-reaching than the applications made in the chemistry of molecules. On the other hand, the possibility to derive particle masses from first principles would seem even more remote, from the standpoint of geometrodynamics, than the possibility to calculate the binding energy of a complex molecule from first principles. The energy with which one hopes to end up, 10^{-27} g to 10^{-24} g, is smaller by twenty powers of ten than the characteristic energy of the theory with which one starts. Still it would seem unwise to discount in advance the ingenuity of available methods to calculate reliably small effects against enormously larger backgrounds, as for example in the case of superconductivity [41] ($\sim 10^{-4}$ eV versus ~ 10 eV).

How can the geometrodynamical interpretation of particles (Table 3) be tested? Sooner to be expected than quantitative calculations are qualitative predictions and conceptual developments. Of such developments none gives more incentive than gravitational collapse [18] to believe in a tie between particles and geometry; and none gives more encouragement to believe in the relevance of the Planck length than the concept of charge as electric lines of force trapped in the topology of space.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

TABLE 3 *Quantum–Geometrodynamical Interpretation of Particles and Forces*

Ultimate dynamical object	Not an electron or any other kind of particle, but geometry itself
Geometry of space	Not unique and classical, but everywhere resonating at the scale of the Planck length between configurations of varied submicroscopic curvature and varied topology
Topology of space	Those geometries that occur with appreciable probability amplitude are highly multiply connected throughout all space ("foamlike structure")
Particle	Not a foreign and physical entity moving about within the geometry of space, but a quantum state of excitation of that geometry itself; as unimportant for the physics of the vacuum as a cloud is unimportant for the physics of the sky. Not a localized ripple in the geometry, not a submicroscopic "wormhole" in the geometry of space at the Planck scale of distances, but an excitonlike change in the phase relations in the probability amplitude for a very large number of such wormholes
Charge	Not a place where Maxwell's equations fail, not a mysterious "foreign and physical" jelly introduced into geometry from outside, but "lines of electric force trapped in the topology of space" (Table 2 and Problem 2)
Spin	Not a dynamical object added to geometry but the nonclassical two-valuedness associated with geometry itself—because distinct probability amplitudes attach to a multiply connected 3-geometry endowed with alternative triad fields or "spin structure" (Problem 2)
Force	Strong forces, weak forces, and intermediate forces no more distinct in their character than van der Waals forces, ionic forces, and valence forces; not themselves primordial, but the residual effect of percentage-wise negligible changes in the enormous density of the energy of the zero-point fluctuations taking place in the geometry of space at small distances

Widening vistas open out for further investigation [42]. (1) What can one give in the way of simple principles to shortcircuit all the usual derivations of Einstein's field equations and pass in one leap from postulates to the Hamilton–Jacobi equation itself? (2) What deeper insights can one win into the structure of superspace? And (3) at what new point in geometrodynamics does one draw that old line between dynamic law and initial conditions that one sees throughout all of physics?

These issues are the foothills. The mountain looms above them: Is a particle a state of excitation of the geometry of space?

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Einstein, above his work and writing, held a long-term vision: There is nothing in the world except curved empty space [43]. Geometry bent one way here describes gravitation. Rippled another way somewhere else it manifests all the qualities of an electromagnetic wave. Excited at still another place, the magic material that is space shows itself as a particle. There is nothing that is foreign and "physical" immersed in space. Everything that is, is constructed out of geometry. This is the dream. Is this dream coming to life?

Note: This article is the revised and extended version of reports made at the Académie Internationale de Philosophie des Sciences, Oberwolfach/ Freiburg im Breisgau, July, 1966; the International School of Nonlinear Mathematics and Physics, Munich, July, 1966; the Society of Engineering Science, Raleigh, North Carolina, October, 1966; the Colloque Internationale sur Fluides et Champ Gravitationel en Relativité Générale, Paris, June, 1967; and the Battelle Rencontres in Mathematics and Physics, Seattle, July, 1967. It appears in the proceedings of the second, third, and fifth organizations in English and in the proceedings of the fourth in French. Appreciation is expressed to the organizers of these conferences for their hospitality and to many colleagues for discussions and advice, among them especially Y. Choquet, Bryce DeWitt, Cecile DeWitt, L. Ehrenpreis, U. Gerlach, H. Leutwyler, C. W. Misner, and M. Stern.

PROBLEM 1. "DERIVATION" OF "EINSTEIN–HAMILTON–JACOBI EQUATION"

If one did not know the Einstein–Hamilton–Jacobi (EHJ) equation (Eq. (12) or (13)), how might one hope to derive it straight off from plausible first principles, without ever going through the formulation of the Einstein field equations themselves? To find one such direct derivation of the EHJ equation would not seem rash to hope for when one already knows five ways to derive the field equations in their traditional form: (1) Einstein's original derivation based upon the correspondence with Newtonian gravitational theory; (2) Weyl's derivation based upon enumeration of all covariant differential operations on the metric tensor that are linear in the second derivatives and that contain no higher derivatives; (3) Hilbert's derivation from a variational principle; (4) Cartan's capture [44] of the geometrical content of the field equations [45]; and (5) the derivation by Gupta, Thirring, and Feynman starting with the theory of a field of spin two and mass zero in flat space.

The central starting point in the proposed derivation would necessarily seem to be "imbeddability." On the basis of experience one can reasonably

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ask that the desired Hamilton–Jacobi equation, of course combined as always with the conditions of constructive interference, should pick out ${}^{(3)}\mathcal{G}$'s that will fit together into a ${}^{(4)}\mathcal{G}$.

In what way would one violate the "condition of imbeddability" if, for example, one left the differential operator unchanged in (13) but replaced the term ${}^{(3)}R$ by the square or by some other function of this curvature scalar? Without directly answering this question, one can say that the special form of the equation is governed in the most direct possible way by the four-dimensional character of space-time. When geometrodynamics is put into simplest terms, it comes out as the statement

$$(\text{CURVATURE}) = (\text{DENSITY OF MASS-ENERGY})$$

In the present considerations one looks apart from situations where there is any "real" mass-energy present. To write the equation as if there were such a term on the right is, however, a reminder that one is speaking about a tensorial quantity, dependent therefore not only upon one's choice of point, but also upon the choice of direction, or unit 4-vector, at that point. This circumstance illuminates the meaning of the curvature term on the left [44, 45]. It has to do with curvature of the 4-geometry in a tangent plane normal to the 4-vector in question. Good! But we are considering conditions, not merely at one point, but at a threefold infinity of points, that form a spacelike 3-geometry. This 3-geometry is not necessarily "free of extrinsic curvature" (vanishing "tensor K_{ij} of extrinsic curvature" or vanishing "second fundamental form") at the point under study. If it is, excellent! Then the desired curvature is given directly by the scalar curvature invariant ${}^{(3)}R$ of the geometry intrinsic to ${}^{(3)}\mathcal{G}$ at the point in question. However, any nonvanishing "extrinsic curvature of the 3-geometry relative to the enveloping 4-geometry" makes an additional contribution to the scalar curvature intrinsic to ${}^{(3)}\mathcal{G}$. One has to correct for this contribution before one secures a proper measure,

$$(\text{Tr } \mathbf{K})^2 - \text{Tr } \mathbf{K}^2 + {}^{(3)}R$$

of the curvature of the 4-geometry itself in the tangent plane in question (Gauss–Codazzi formula). Why this special bilinear expression in the tensor K_{ij} of extrinsic curvature and not some other one [46]? The rationale shows most directly when one considers the special case where the 4-geometry itself is flat. In this case the "correction terms" in the expression for the desired curvature must exactly compensate ${}^{(3)}R$. Let the 3-geometry be imbedded in the 4-geometry with "principal radii of curvature" at the point in question equal to ρ_1, ρ_2, and ρ_3. Then the scalar curvature invariant of the intrinsic geometry is ${}^{(3)}R = -2/\rho_2\rho_3 - 2/\rho_3\rho_1 - 2/\rho_1\rho_2$. The difference in sign here compared to the familiar ${}^{(3)}R = 6/a^2$ for a 3-sphere of radius a arises from the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

fact that the "radius of curvature" is being measured in a 4-geometry of signature $- + + +$. The tensor of extrinsic curvature is

$$\mathbf{K} = \begin{Vmatrix} \frac{1}{\rho_1} & 0 & 0 \\ 0 & \frac{1}{\rho_2} & 0 \\ 0 & 0 & \frac{1}{\rho_3} \end{Vmatrix}$$

It is desired to build up out of this tensor an expression that (1) is bilinear in the reciprocals of the ρ_i; (2) contains no terms in $1/\rho_1^2$, etc.; and (3) compensates ${}^{(3)}R$. These three requirements lead uniquely to the stated formula. The considerations outlined here by no means come close to providing a derivation of the Einstein–Hamilton–Jacobi equation, but they perhaps suggest some of the factors that may play a natural part in such a derivation.

Subscript on Relation of Hamilton–Jacobi Method to Conventional Analytic Solutions of Field Equations

To look toward the Hamilton–Jacobi equation as an illuminating way to found general relativity is not to look away from the field equations as a way to solve problems in general relativity. The Hamilton–Jacobi equation is divorced from the equations of motion no more in geometrodynamics than in particle mechanics. Even in the most elementary problem the connection between the two methods is inescapable. Yes, one can solve Lagrange's equation

$$m\frac{d^2x}{dt^2} + \frac{\partial V}{\partial x} = 0$$

by direct numerical integration. Yes, one can translate the problem into Hamilton–Jacobi formalism, write down the equation

$$-\frac{\partial S}{\partial t} = \left(\frac{1}{2m}\right)\left(\frac{\partial S}{\partial x}\right)^2 + V(x, t)$$

and integrate *it* by the most elementary numerical methods (replacement of (x, t) continuum by lattice space, $t = m\tau$, $x = n\delta$, with m and n integers; replacement of partial differential equation by difference equation). However, the most penetrating method to integrate the partial differential equation has long been known to go straight back to the Lagrange equation itself for its

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

start. Not everywhere does one at first evaluate $S(x, t)$, but only along a classical world line or history H,

$$x = x_H(t)$$

$$\dot{x} = \frac{dx_H(t)}{dt}$$

thus

$$s(t) = S(x_H(t), t) = \int_{t_0}^{t} [\tfrac{1}{2}m\dot{x}_H^2 - V(x_H(t), t)]\, dt$$

or, in a more general problem, with Lagrange function L,

$$s(t) = S(x_H(t), t) = \int_{t_0}^{t} L(\dot{x}_H, x_H, t)\, dt$$

From a knowledge of S along the world line one goes to a knowledge of S in a narrow band on either side of the world line by using the relations

$$\begin{pmatrix}\text{RATE OF CHANGE}\\ \text{OF ACTION } S\\ \text{WITH POSITION } x\end{pmatrix} = (\text{MOMENTUM}) = \frac{\partial L}{\partial \dot{x}}\,(\text{GENERAL}) = m\dot{x}_H(t)\,(\text{HERE})$$

and

$$-\begin{pmatrix}\text{RATE OF CHANGE OF}\\ \text{ACTION } S \text{ WITH}\\ \text{TIME } t\end{pmatrix} = (\text{ENERGY}) = \dot{x}\frac{\partial L}{\partial \dot{x}} - L\,(\text{GENERAL})$$

$$= \tfrac{1}{2}m\dot{x}_H^2 + V(x_H(t), t)\,(\text{HERE})$$

Thus at the point

$$x = x_H(t^*) + \delta x$$

$$t = t^* + \delta t$$

a little way off the world line, the Hamilton–Jacobi function ("classical phase"; $\hbar$ times phase of quantum-mechanical wave function) is

$$S(x, t) = s(t^*) + \left(\frac{\partial S}{\partial x}\right)\delta x + \left(\frac{\partial S}{\partial t}\right)\delta t$$

An identical procedure gives $S_{\text{new}}(x, t)$, a new solution with infinitesimally different initial conditions, throughout a band extending on either side of an infinitesimally different world line, $S_{\text{new}}(x, t)$. The "condition of constructive interference" between the new "wave" and the old "wave,"

$$S(x, t) = S_{\text{new}}(x, t)$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

gives back at once the original solution of the equation of motion—and all the history H that goes with that solution. So too in geometrodynamics!

One can use a known solution of the Einstein field equations to find the Hamilton–Jacobi function $S({}^{(3)}\mathcal{G})$ throughout a certain narrow swathe through superspace. For this purpose one will find it easiest to start with one of the well-known analytic solutions of Einstein's field equations. Let the 4-geometry be written in the form

$$ds^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta$$

where the metric coefficients $g_{\alpha\beta}$ are certain known functions of the four coordinates x^μ. The 4-geometry expressed in this way summarizes the dynamical history H of the 3-geometry of space. It also defines a submanifold—call it H—of the superspace $\mathcal{S}$. How now to determine the values taken on by the Hamilton–Jacobi function throughout the submanifold? First define the equivalent of the starting time t_0 in the one-particle problem. This equivalent is not a single number t_0, but in general a different number $t_0(x, y, z)$ for each point in 3-space; or better stated, it is a spacelike initial value hypersurface σ_0 slicing through the given ${}^{(4)}\mathcal{G}$. Next, give the equivalent of the point $x_H(t_0)$, t_0 along the classical history. It is the "momentary state of the geometry of space" on the hypersurface σ_0; that is, it is the 3-geometry ${}^{(3)}\mathcal{G}_0$ defined by the metric

$$ds^2(\sigma_0) = \left[g_{00}\left(\frac{\partial x^0}{\partial x^m}\right)\left(\frac{\partial x^0}{\partial x^n}\right) + 2g_{0m}\left(\frac{\partial x^0}{\partial x^n}\right) + g_{mn} \right] dx^m\, dx^n$$

Similarly with the equivalent of the running time t and the state $x_H(t)$ of the particle at this time. Thus, give a spacelike hypersurface σ by giving $t = t(x, y, z)$; and calculate the corresponding metric $ds^2(\sigma)$, which defines a 3-geometry ${}^{(3)}\mathcal{G}$ "on the classical history H." The value of the classical phase function S on this classical history is given by a fourfold integral. This integral is extended over the region of spacetime bounded by the two hypersurfaces. It has the form

$$s(\sigma) = \int_{\sigma_0}^{\sigma} \mathcal{L}(x^\alpha)\, d^4x$$

Here the Lagrange density $\mathcal{L}$ is given, for example, by Arnowitt, Deser, and Misner ([2]; see also GMD). Consider now a "point" ${}^{(3)}\mathcal{G}$ in superspace a little ways removed from some "point" ${}^{(3)}\mathcal{G}^*$ that lies on the classical history H. What is the value of S for this new 3-geometry? The difference between the two 3-geometries is expressed most conveniently for the present expository purpose in terms of the difference of the metric coefficients at corresponding points:

$$\delta g_{mn}(x, y, z) = g_{mn}(x, y, z; \sigma) - g_{mn}(x, y, z; \sigma^*)$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

This difference in 3-geometries has to be multiplied by the value of the conjugate geometrodynamical momentum $\pi^{mn}(x, y, z)$ at the point ${}^{(3)}\mathcal{G}^*$ on the classical history H (details of definition and calculation of π^{mn} given, for example, in GMD) and integrated to give the change in the Hamilton–Jacobi function. Thus, throughout a thin swathe in superspace one can write

$$S({}^{(3)}\mathcal{G}) = s(\sigma^*) + \int \pi^{mn} \delta g_{mn}\, d^3x$$

PROBLEM 2. STRUCTURE OF SUPERSPACE

Spacetime is the arena of particle dynamics. Superspace is the arena of geometrodynamics. How different our knowledge of these two arenas! In particle dynamics the structure of spacetime is taken to be given as from on high: an everywhere flat ideal Minkowski–Lorentz manifold. In geometrodynamics the structure of superspace is to be considered as defined entirely internally; that is to say, by the very form of the "Einstein–Schroedinger equation" itself. We write this equation symbolically as

$$-\frac{\nabla^2 \psi}{(\delta^{(3)}\mathcal{G})^2} + {}^{(3)}R\psi = 0$$

To ask about the structure of this equation is therefore nothing more or less than to ask, what is the structure of superspace?

It is conceivable that a given geometry may be more appropriately identified with a whole class of points in superspace than with a single point, when that 3-geometry has less than maximal symmetry. In this way, Professor Stephen Smale kindly points out, one may keep from introducing "conical singularities" in superspace. To avoid such singularities would seem to be essential if superspace is to rise from a mere topological manifold to the status of a differentiable manifold.

Quantum theory once known, classical theory follows easily and uniquely, by way of the correspondence principle,

$$\psi \simeq \begin{pmatrix}\text{SLOWLY VARYING} \\ \text{AMPLITUDE FACTOR}\end{pmatrix} \exp\left(\frac{iS}{\hbar}\right)$$

However, classical theory alone being known, it is ordinarily difficult or impossible uniquely to determine the form of the quantum wave equation [47]. No system better illustrates this point than a particle moving in a prescribed external electromagnetic field. To know the Hamilton–Jacobi equation alone

$$g_{\alpha\beta}\left(\frac{\partial S}{\partial x^\alpha} + \frac{eA_\alpha}{c}\right)\left(\frac{\partial S}{\partial x^\beta} + \frac{eA_\beta}{c}\right) + m^2c^2 = 0$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

is to have no way to decide between the Schroedinger–Klein–Gordon equation, the Dirac equation, or one or another wave equation of higher spin; they all reduce to the same Hamilton–Jacobi equation in the appropriate semi-classical limit. Similarly in geometrodynamics. From the well-defined "Einstein–Hamilton–Jacobi" equation (12, 13)

$$\left(\frac{\nabla S}{\delta^{(3)}\mathcal{G}}\right)^2 + {}^{(3)}R = 0$$

no one knows a satisfying way to go to a unique "Einstein–Schroedinger" equation [48].

In the example of the particle most of the ambiguity about the form of the wave equation is resolved as soon as observation or other evidence reveals the spin [49]—or more directly—how many possible orientations there are for any spin degree of freedom over and above the three degrees of freedom that tell location in space. In other words, in addition to the Hamilton–Jacobi equation, one must know the "structure of configuration space" in order to end up with a well-defined wave equation. Hence our question here: What is the structure of superspace?

Unravel the structure of superspace? Hardly all in one jump, and hardly today! Instead a step by step penetration of the issues is more to be anticipated, if the history of electromagnetism or atomic structure or other branches of physics is any guide. At least three levels of analysis are to be perceived. First, one already knows the structure relevant to classical geometrodynamics: (a) Superspace has Hausdorf topology (Stern [4]). (b) Superspace is endowed with an indefinite metric, defined by the expression [50]

$$\left(\frac{1}{2g}\right)(g_{ik}g_{jl} + g_{il}g_{jk} - g_{ij}g_{kl})$$

that occurs in the Hamilton–Jacobi equation (Eq. (12)). Second, one searches for those features in the structure of superspace that come into play when space changes its topology. Third, one hopes in the longer term to uncover those deeper properties of superspace that give ordinary space its dimensionality, its metric structure, and its ability to propagate electromagnetic fields and neutrinos with the speed of light. It is appropriate to discuss the structure of superspace a little further at all three of these levels.

Level 1. Classical Geometrodynamics; Topology Does Not Change

In classical geometrodynamics space does not change its topology. Space may be "preparing" to change its topology, but it cannot actually make the change within the context of classical theory [51]. It can at most signal its "intention" to change topology by developing somewhere a curva-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ture that increases without limit ("gravitational collapse" [18]). To go further with the analysis of the collapse phenomenon and treat changes in topology forces one to go outside the framework of classical theory. So long as one stays inside classical theory, he must restrict his attention in one dynamical problem to one topology. Which topologies are acceptable?

Of all topologies none is more familiar than that associated in Einstein's theory with the geometry in and around a dilute center of attraction such as the sun. Curved at small distances, the geometry becomes asymptotically flat at large distances. The topology, as distinguished from the geometry, is Euclidean: E_3 or $R \times R \times R$. In a broader context, however, Einstein thought of the space around one star as part of the space engulfing all stars [52], and of this total space as closed and endowed with the topology of the 3-sphere, S_3. No one has stated more strongly than he the arguments for considering space to be closed [53].

One acquires a new reason to consider space to be closed when he considers how hard it is to define "an open and asymptotically flat space" in the context of quantum geometrodynamics: Those ${}^{(3)}\mathscr{G}$'s that occur with overwhelming probability are everywhere endowed with all kinds of ripples and other geometrical structures at the scale of the Planck length. There is no direction that one can take and there is no distance that one can travel that will erase this structure. Under these circumstances it is difficult to attribute any well-defined meaning whatsoever to the term "asymptotically flat." On the other hand, one knows more than one example of a space that is open and *not* asymptotically flat, and that becomes wilder and wilder at great distances [54]. Not having any means to distinguish one open space from another as "good," one would seem justified at this stage in the development of the subject to exclude from attention all open spaces. This approach is the more attractive in that it appears possible [55] to distinguish between one type of closed 3-geometry and another by straightforward classificatory integers similar to the Betti numbers. In contrast, the concept of "asymptotically flat"—even in contexts where it is relevant—is much more complex to formulate [56].

Out of a manifold with the topology E_3 it is possible to cut out a block with the shape of a cube and obtain the topology of the 3-torus, $S_1 \times S_1 \times S_1$. On a manifold with this topology the initial value problem of classical geometrodynamics presents difficulties in certain cases, according to unpublished considerations of Brill and Avez. It is conceivable that these difficulties indicate that the topology $S_1 \times S_1 \times S_1$ is not acceptable. In that event one might almost say that this topology has inherited a "defective gene" from its parent topology E_3.

Another "black gene" that one can perhaps reasonably exclude is nonorientability. We assume in effect that "transport" of a right-handed glove shall never bring it back to its starting point left handed.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Compatible with these very tentative principles for selecting "acceptable topologies" are the 3-sphere (S_3); the 3-sphere with addition of one handle or "wormhole" ($S_2 \times S_1 = W_1$); and the 3-sphere with n wormholes (W_n). There may or may not be further acceptable topologies not included in this list [55].

Fixed Topology Excludes GMD Account of Particles and Fields

Taking the most familiar of these "acceptable topologies," S_3, as case example, one can discuss and treat quantitatively at the classical level a rich variety of physical processes, including gravitational radiation [38]; gravitational geons [20]; the planetary motion, collisions, and breakup of geons [20]; the Taub model of an expanding and recontracting universe [57]; and more complex model universes [38]. Even so, one is limited in the physics that he can include in any of these models. One has no place for particles. Nor can one treat classically the final stages of gravitational collapse. These limitations seem unrelated to each other. In classical theory they are unrelated. In that view, the geometry of space remains forever tied to a unique topology. Not so in the purely quantum-geometrodynamical model of physics. There (1) a particle is pictured in terms of space resonating from one topology to another [58]; and there (2) the final stages of gravitational collapse are viewed as a coupling of macroscopic motion and microscopic topology ("waterfall and foam"). In excluding from consideration all changes in topology, classical theory, on these views, also excludes any account of the rationale of particles. Particles have to be introduced as foreign and physical entities and space has to be viewed as arena, not as structural material. In this limited conceptual framework the particles and other fields are counted as having degrees of freedom over and above those of the geometry. The electromagnetic field, in particular, is viewed as an entity additional to geometry.

Attempts have been made to regard electromagnetism as an aspect of geometry. This enterprise limits itself to that classical framework of ideas where one looks apart from submicroscopic fluctuations in the geometry and the topology of space. One attempt has had some minor success. The second-order equations of Maxwell and the second-order equations of Einstein have been combined into one set of equations of the fourth order that make no reference to any but geometric magnitudes [59]. The Maxwell field is recognized by the "footprints" it leaves on the geometry of space. In a certain sense those "footprints" *are* the electromagnetic field; hence the name, "already unified theory" of gravitation and electromagnetism. Faithfully though this theory reproduces in other respects the dynamic content of the equations of Maxwell and Einstein, it turns out not to be adapted to treat the initial value problem. Geometrical measurements alone on an initial space-like hypersurface do not always suffice completely to determine the future time

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

evolution of the geometry [60]. In other words, one cannot uphold in "already unified field theory" a strict division of dynamics into "equations of motion" and "initial value data for these equations." Yet on that division one had learned to insist, not least by reason of hard-won lessons from the Hamilton–Jacobi theory and from quantum theory. Therefore the electromagnetic field, like a particle, can hardly be treated as anything but a foreign or "physical" entity immersed in space so long as one looks away from the microscopic geometry of space.

Formalism of Field When Treated as "Foreign and Physical"

In this nongeometric or "arena" approach to physics one considers the degrees of freedom of the "foreign and physical" entity as additional to those of the geometry. In the most elementary example, the case where only one such entity is contemplated, the pure source-free electromagnetic field, one writes the Hamilton–Jacobi function or the Schroedinger probability amplitude, as the case may be, in the form

$$S = S(g_{ik}(x, y, z), A_m(x, y, z))$$

or

$$\psi = \psi(g_{ik}, A_m)$$

Similarly when more fields are involved. Even at this level of analysis the geometrical approach has a contribution to make. The Hamilton–Jacobi function, ostensibly dependent upon the individual components of the metric and of the electromagnetic vector potential, actually has to be understood to depend only on the 3-geometry ${}^{(3)}\mathcal{G}$ and on the 2-form $\mathbf{B} = \mathbf{dA}$. Consequently it cannot change S to "push the rubber sheet on which the coordinates are painted" over space in such a way that the point P, which formerly had the coordinates x^i, now acquires the coordinates $x^i - \xi^i$. Similarly the point $P + dP$, which formerly had the coordinates $x^i + dx^i$, now acquires the coordinates

$$x^i + dx^i - \xi^i - \left(\frac{\partial \xi^i}{\partial x^j}\right) dx^j$$

The separation between the two points of course remains unchanged by this alteration in coordinates:

$$\begin{aligned} ds^2 &= g_{ij}(x^s)\, dx^i\, dx^j \\ &= g_{ij}^{\text{new}}(x^s - \xi^s)(dx^i - \xi^i_{,m}\, dx^m)(dx^j - \xi^j_{,n}\, dx^n) \end{aligned}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

To terms of the first order in the displacement ξ^s the alteration in the metric coefficients is

$$\delta g_{ij} = g_{ij}^{\text{new}} - g_{ij} = \left(\frac{\partial g_{ij}}{\partial x^s}\right)\xi^s + \xi_{i,j} + \xi_{j,i} = \xi_{i\,|\,j} + \xi_{j\,|\,i}$$

where the subscript $|\,j$ denotes covariant differentiation, in the space of the 3-geometry, with respect to the jth coordinate. In a similar way one finds the alteration in the electromagnetic vector potential,

$$\delta A_i = A_i^{\text{new}} - A_i = A_{i,s}\xi^s + A_j \xi^j_{,i}$$

These changes in the metric coefficients, taking place throughout all space, ostensibly produce a change in the Hamilton–Jacobi function determined by the functional derivatives of S; thus,

$$\delta S = \int \left[\left(\frac{\delta S}{\delta g_{ij}}\right)\delta g_{ij} + \left(\frac{\delta S}{\delta A_i}\right)\delta A_i\right] d^3x$$

However, the mere shift in coordinates cannot produce any real physical change. Consequently the quantity δS must vanish, independent of the choice of the coordinate displacements ξ^i. The consequences of this coordinate invariance are easily traced out. The functional derivatives that come into consideration have well-defined physical interpretations. Thus

$$\pi^{ij} = \frac{\delta S}{\delta g_{ij}}$$

is the geometrodynamical momentum [2, 3] conjugate to the geometrodynamical field coordinate g_{ij}. It is closely connected with the "extrinsic curvature" K^{ij} of the 3-geometry with respect to the yet-to-be constructed 4-manifold. The other derivative,

$$\mathbb{E}^i = \frac{\delta S}{\delta A_i}$$

the electrodynamic momentum conjugate to the vector potential A_i, is nothing other than the electric field. This quantity satisfies the divergence condition [61]

$$\mathbb{E}^i_{,i} = 0$$

The condition that S be invariant with respect to coordinate changes takes the form

$$0 = \delta S = \int [\pi^{ij}(\xi_{i\,|\,j} + \xi_{j\,|\,i}) + \mathbb{E}^i(A_{i,j}\xi^j + A_j \xi^j_{,i})]\, d^3x$$

Integrating by parts, readjusting the positions of indices as appropriate, and making use of the divergence relation, one finds

$$0 = \int [-2\pi^s_{js|} + \mathbb{E}^i(A_{i,j} - A_{j,i})]\xi^j\, d^3x$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

This must vanish for arbitrary choices of the field of coordinate displacements ξ^j. Consequently the quantity in square brackets must vanish. One finds in this way those three of Einstein's field equations that link the curvature of space with the Poynting density of flow of electromagnetic field energy [62]:

$$2\pi T^{\perp s}_{js} = (\mathfrak{E} \times \mathbf{B})_j$$

How remarkable that one gets so much—including the Poynting vector itself—from the elementary condition that S should depend, not upon the components g_{ij} and A_i individually, but only upon the coordinate-independent geometrical quantities that are "dressed up" in these components [63]!

Express electrodynamics in geometric language, yes. See the footprints of the electromagnetic field on the geometry of space, yes. Conceive of classical geometrodynamical electric charge as lines of force trapped in the topology of space, yes. But explain geometrically the "necessity" for electromagnetism and other "physical" entities, no. Not within the context of classical geometrodynamics and its never-changing topology.

Level 2. Space Resonating between 3-Geometries of Varied Topology

A new world opens out for analysis in quantum geometrodynamics. The central new concept is space resonating between one foamlike structure and another. For this multiple connectedness of space at submicroscopic distances no single feature of nature speaks more powerfully than electric charge. Yet at least as impressive as charge is the prevalence of spin $\frac{1}{2}$ throughout the world of elementary particle physics. "It is impossible to accept any description of elementary particles that does not have a place for spin $\frac{1}{2}$." This quotation from the book *Geometrodynamics* [64] of 1962 goes on to say, "What then has any purely geometrical description to offer in explanation of spin $\frac{1}{2}$ in general? More particularly and more importantly, what possible place is there in quantum geometrodynamics for the neutrino—the only entity of half integral spin which is a pure field in its own right, in the sense that it has zero rest mass and moves with the speed of light? No clear or satisfactory answer is known to this question today. Unless and until an answer is forthcoming, *pure quantum geometrodynamics must be judged deficient as a basis for elementary particle physics.*" Happily the concept of spin manifold [65] has subsequently come to light, not least through the work of John Milnor. This concept suggests a new and interesting *interpretation of a spinor field* within the context of the resonating microtopology of quantum geometrodynamics, *as the nonclassical two-valuedness that attaches to the probability amplitude for otherwise identical 3-geometries endowed with alternative "spin structures."*

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

The Orientation Entanglement Relation or "Version"

Spin and topology: what is the connection? Take a cube (Fig. 7). To its upper northeast corner attach one end of a long elastic string. Run the other end to the upper northeast corner of the room and fasten it there. With seven other elastic strings attach the other corners of the cube to the corresponding corners of the room. Now select any axis running through the center of the cube and rotate the figure about that axis through 360°. The cube resumes its original configuration. Not so the strings. They are in a

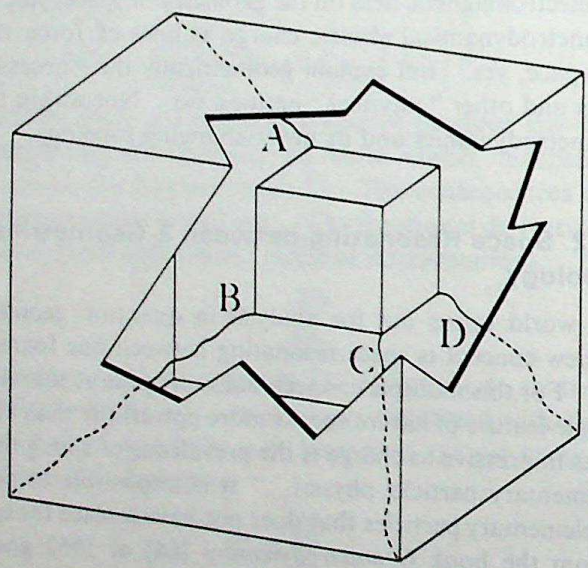


FIGURE 7. The elastic strings attached to the central cube keep account of its orientation entanglement relation with its surroundings.

tangle. Moreover, rejecting cuts, one has no way to untangle the strings. Consequently it needs more than orientation to tell the relation between the cube and its surroundings. The necessary bookkeeping is provided by a spinor,

$$s = \begin{pmatrix} \xi \\ \eta \end{pmatrix}$$

Under a rotation through the angle θ about an axis making angles α, β, γ with the x, y, and z axes this spinor is transformed to the new spinor $s' = \mathbf{q}s$. Here

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the quaternion or "versor" $\mathbf{q}$ has a value indicated in various ways in various well-known systems of nomenclature:

$$\mathbf{q} = \cos \tfrac{1}{2}\theta + \sin \tfrac{1}{2}\theta(\mathbf{i}\cos\alpha + \mathbf{j}\cos\beta + \mathbf{k}\cos\gamma)$$

$$(\text{with } \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \text{ etc.})$$

$$= \cos \tfrac{1}{2}\theta - i \sin \tfrac{1}{2}\theta(\sigma_x \cos\alpha + \sigma_y \cos\beta + \sigma_z \cos\gamma)$$

$$(\text{with } i = (-1)^{1/2} \text{ and } \sigma_x\sigma_y = -\sigma_y\sigma_x = i\sigma_z, \text{ etc.})$$

$$= \begin{Vmatrix} (\cos \tfrac{1}{2}\theta + i \sin \tfrac{1}{2}\theta \cos\gamma) & \sin \tfrac{1}{2}\theta(\cos\beta - i\cos\alpha) \\ \sin \tfrac{1}{2}\theta(-\cos\beta - i\cos\alpha) & (\cos \tfrac{1}{2}\theta - i \sin \tfrac{1}{2}\theta \cos\gamma) \end{Vmatrix}$$

The important point is the change of sign under a 360° rotation: $\mathbf{q}(360°) \cdot s = -s$. The 360° rotation alters what one may most appropriately call the "orientation entanglement relation" between the cube and its surroundings—or, more briefly, the "version" of the cube. The spinor keeps account of this orientation entanglement relation. Two successive rotations by 360° restore the cube to its original orientation entanglement relation with its surroundings. The strings at first appear to be tangled up with twice the twist they had before. Nevertheless, they can now be untangled completely, as one confirms by direct trial or by elementary reasoning [66, 67].

Arbitrarily pick out one way of placing the cube, call it the "standard orientation entanglement relation" between the cube and its surroundings, or the standard "version" of the cube, and associate with it the spinor $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Proceed similarly at other points of space, taking care only that any changes in the standard version from point to point shall take place smoothly.

A special situation develops when an orientable 3-geometry [68] is endowed with a handle or wormhole. The cube can be transported in imagination from A to B "through the surrounding nearby space" or "through the wormhole." With one cube possessing a certain pattern of colored faces follow one route and with another identically colored cube follow the other route. It makes physical sense to ask if the two cubes have at B identical orientation entanglement relations to their surroundings [67]. If they do not, alter the definition of the standard orientation within the wormhole by a rotation. Let this rotation increase continuously from zero at the mouth of the handle near A to 360° at the mouth near B. At each point outside the wormhole a triad of axes a, b, c defines the direction of the axes of the cube when the cube is located at that point in its standard orientation. These triads are not affected by the alterations made inside the wormhole. They vary in direction as smoothly as ever from point to point. At the two mouths of the wormhole they join on as smoothly to the new field of triads inside as they did to the old field of triads. Now, however, one has at last

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

achieved an everywhere continuous field of standard orientation entanglement relations. The associated spinor field is likewise continuous, having everywhere the standard value $\begin{pmatrix}0\\1\end{pmatrix}$, whereas before it underwent somewhere a discontinuous change from this value to its negative. One can restate in mathematical language what has been accomplished. First, one has laid down a "spin structure" upon the manifold. This structure is defined by the field of triads, or by "the class of such fields that are equivalent under homotopy." Second, one has laid down a "spinor field" upon the manifold. This spinor field is defined with respect to the given "spin structure." Other spinor fields can of course be laid down, with components also varying continuously from place to place.

Alternative Spin Structures in a Multiply Connected Space

Start again with another closed orientable 3-manifold endowed with the original topology and metric. One might hope to obtain an acceptable spin structure on this new manifold by taking over to it the identical field of triads that serves for the original manifold. It may be that this supposition is correct about the particular new manifold that one happens to have picked out. In this event identical colored cubes "taken from A to B by inequivalent routes" will preserve their orientation entanglement relation, one to the other. However, the supposition can equally well be incorrect. If so, the consequences are direct. The two cubes, carried by different routes from A to B, always in alignment with the canonical field of triads, will end up at B inequivalent to the extent of a 360° rotation. Moreover, this inequivalence in the orientation entanglement relation with the surroundings can be detected in principle by direct physical measurement [67]. In other words, the spin structure of the original manifold does not apply to the new manifold. To serve in the new manifold, the field of triads has to be modified to the extent of a 360° rotation within the wormhole or by an equivalent change. The difference between the new manifold and the original manifold expresses itself in these terms: the two manifolds have the same topology and metric, but they have inequivalent spin structures. *The difference between the two manifolds is not only mathematical. It is physical.*

The Multisheeted Character of Superspace

One does not classify the closed orientable 3-manifold of physics completely when he gives its topology, its differential structure, and its metric. He must tell in addition which spin structure it has. The spin structure, like the metric, lends itself in principle to observation. Consequently it is not enough for the purposes of quantum geometrodynamics to give the probability

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

amplitude for a "3-geometry" as the term 3-geometry was previously understood. One must introduce a new two-valued descriptor, w_k, for each ($k = 1, 2, \ldots, n$) of the n wormholes of the manifold, to distinguish the two inequivalent ways to lay down a spin structure in the interior of that wormhole. The new, enlarged concept of a 3-geometry ${}^{(3)}\mathcal{G}$ adjoins these descriptors to the continuous infinity of parameters which alone served previously to distinguish one 3-geometry, ${}^{(3)}\mathcal{G}^{\text{old}}$, from another; thus,

$$ {}^{(3)}\mathcal{G} = ({}^{(3)}\mathcal{G}^{\text{old}};\ w_1, w_2, \ldots, w_n) $$

and

$$ \psi({}^{(3)}\mathcal{G}) = \psi({}^{(3)}\mathcal{G}^{\text{old}};\ w_1, w_2, \ldots, w_n) $$

In other words, *superspace acquires a multisheeted character* (Fig. 8) with 2^n distinct sheets in that region of superspace where the 3-geometry is endowed with n wormholes.

For the two values to be assigned to the descriptor w_n it is natural to pick $+1$ and -1. However, there is nothing anomalous about the one spin structure as compared to the other. No canonical way has ever been proposed to give preference to one as compared to the other. Therefore it is a matter of arbitrary choice to which spin structure to assign the descriptor $+1$, and to which, -1.

The words "spin structure" can mislead. They suggest that there is something special about "laying down a spinor field" upon the manifold. They conjure up visions of laying down upon the 3-geometry other kinds of fields that transform according to groups other than the spinor group $SU(2)$; for example, $SU(n)$ or $SL(n)$. However, the relevant point in the whole analysis is not the field of spinors, but the field of triads and their orientation entanglement relations. There is nothing in the concept of spin structure that one could not have conveyed, with less chance of being misunderstood, by using the phrase "triad structure." Moreover, there is not the slightest indication that there is any other structure of a closed orientable 3-manifold that remains to be brought to light. Consequently we take ${}^{(3)}\mathcal{G}$ in its new and enlarged sense to be the full indicator of the configuration of space, and as containing the full set of variables upon which ψ depends. In other words, we take the new ${}^{(3)}\mathcal{G}$ to comprise a set of commuting observables, *complete* in the sense of quantum mechanics, and therefore suitable for analysis of the probability amplitude ψ.

We do not add a spinor field to geometry. Quite the contrary. We take a spinor field away from geometry. A 3-geometry, augmented by a spin structure (as might for example be indicated by the descriptor $(w_1, w_2, \ldots, w_5) = (+1, +1, +1, -1, +1)$) is a possible habitation for a spinor field—but we have thrown out the inhabitant [69].

Spin $\frac{1}{2}$, if it occurs naturally in the context of quantum geometrodynamics, can hardly occur in any other sense than that in which Pauli spoke of

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

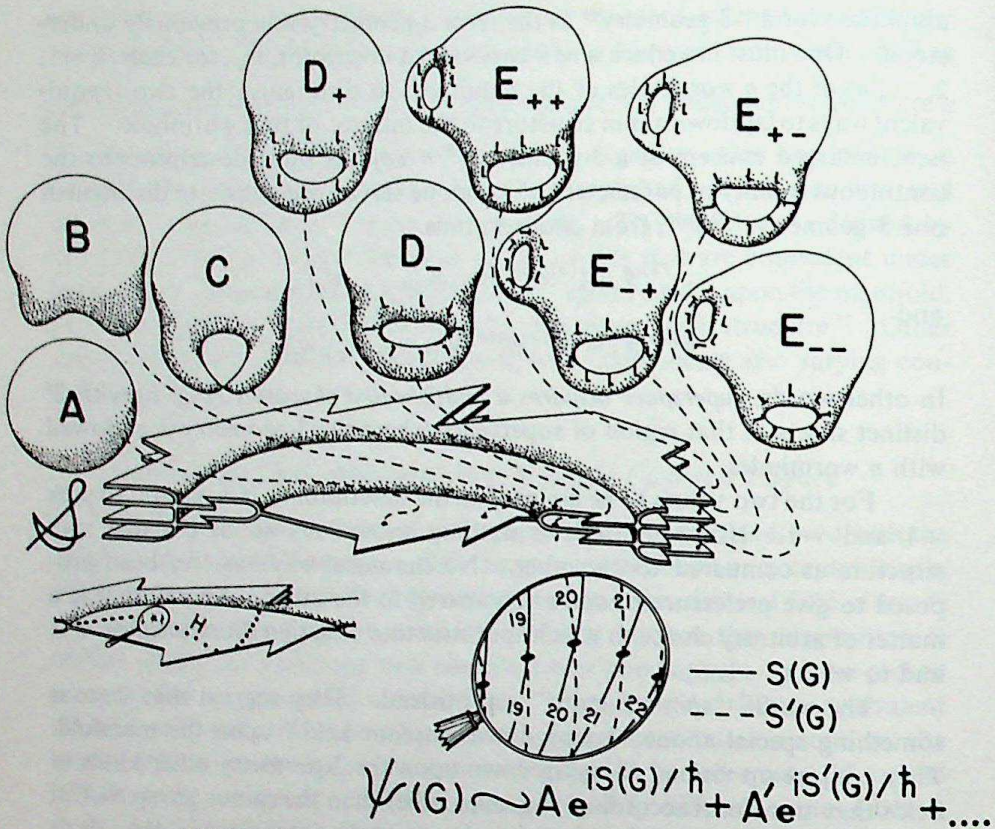


FIGURE 8. The multisheeted character of superspace.

spin from the very beginning, as a "nonclassical two-valuedness." Moreover, a nonclassical two-valuedness is already inescapable in the formalism. There are separate probability amplitudes for a 3-geometry with descriptor $w_k = +1$ and for an otherwise identical 3-geometry with descriptor $w_k = -1$. Does this circumstance imply that quantum geometrodynamics supplies all the machinery one needs to describe fields of spin $\frac{1}{2}$ in general and the neutrino field in particular? That is the proposal. That is the only way that has ever turned up within the framework of Einstein's general relativity and Planck's quantum principle to account for spin. Is this the right path? It is difficult to name any question more decisive than this in one's assessment of "everything as geometry."

Electromagnetism as a Statistical Aspect of Geometry? Other Questions

It would be tempting to stop with this major issue if it did not open the door to so many tributary questions. (1) When a new handle develops and the number of descriptors rises by one, what boundary condition in super-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

space connects the probability amplitude ψ for 3-geometries of the original topology with the probability amplitudes ψ_+ and ψ_- for the two spin structures of the new topology? (2) One may wish to describe for each newly developing wormhole something like "the fractional amplitude going from ψ into ψ_+." However, with the fantastic number of wormholes that typically come into consideration ($\sim 10^{99}/\text{cm}^3$) any such individual bookkeeping would seem for many purposes to be out of the question. It is equally impractical to keep account of the orientation of each of the $\sim 10^{23}$ spins in a ferromagnet. It is much more appropriate to speak of the "density of magnetization" and of perturbations in that density carried by "magnons" [71]. What are the analogous quantities and concepts in geometrodynamics? What statistical approach is best suited to keep account of the many "matching ratios" ψ_+/ψ associated with all the nascent wormholes in the geometry? (3) One speaks of "magnetization," knowing well that the term "magnetization" has not the slightest real meaning at subatomic distances. Is the term "electromagnetic field" equally without submicroscopic physical significance? In other words, among the statistical parameters most appropriate for keeping account of the 10^{99} "matching ratios" per cubic centimeter, *is there one set of statistical parameters that one can identify with the electromagnetic field?*

The Example of Two-Geometries

No one can ask about the physics of changes in the topology of 3-space without at least a look at the mathematics of changes in the topology of 2-space. The superspace built on all 3-geometries is like no mathematical object so much as the superspace "built on all 2-geometries." No one did so much to bring this mathematical object to light as Riemann, the same Bernhard Riemann who taught that the curvature of space is a branch of physics, and who provided the mathematical machinery to describe not only curvature (the Riemann curvature tensor, $R_{\alpha\beta\gamma\delta}$) but also topology (the Betti numbers R_n).

In his 1857 paper Riemann noted that all algebraic 2-geometries endowed with the topology of a 2-sphere are equivalent to each other under conformal transformation (multiplication of all three metric coefficients by a common position-dependent factor λ). In other words the equivalence class of conformally equivalent 2-geometries of the given topology (S_2; or "genus $g = 0$") consists of a single object. It therefore constitutes a single "point" in what is not really a superspace itself, as we have been using that term, but a "reduced superspace": reduced in the sense that the "$(\infty^2)^\infty$" degrees of freedom in λ have here been "strained out" of superspace.

Two-geometries with the topology of the torus (T_2; or one wormhole W_1; or "genus $g = 1$"), Riemann showed, are not all equivalent to one another under conformal transformation. Instead, when conformally

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

equivalent 2-geometries of the topology T_2 are identified, the family of objects that results is a complex continuum of dimension 1 (two real dimensions) [72]. Two-geometries endowed with a larger number of wormholes (topology W_g; genus $g \geqslant 2$), after extraction of the conformal degrees of freedom, reduce to a family of objects described by $3g - 3$ complex parameters ($6g - 6$ real parameters). "Reduced superspace," built on the totality of conformally equivalent closed orientable 2-geometries of all topologies, thus appears to consist of a series of disjoint spaces, the first of dimension 0, the second of dimension 1, the next of dimension 3, the next of dimension 6, and so on. However, great developments in the analysis have taken place since the days of Riemann through the efforts of many investigators [73]. Today, thanks not least to the works of Lipman Bers, one knows how to define one single infinite-dimensional reduced superspace in which all these apparently disparate parts fit smoothly together [74]. What a model for the mathematical treatment of the superspace of general relativity! Yet, at the risk of seeming overdemanding, one has to ask for more. The superspace of physics is not to have any "conformal factor strained out of it"; rather, it is if anything to be enlarged, so as to include all the descriptors w_k of the spin structure. How fit all the pieces of *this* superspace smoothly together? Challenging problem, at the very heart of quantum geometrodynamics!

Other Aspects of Superspace

Why all this emphasis on the structure of superspace? Why not simply spell out explicitly the form of the "Einstein–Schroedinger equation"? For the more elementary problem of a particle moving in flat 3-space it was straightforward for Schroedinger to derive his wave equation. Simple considerations of invariance with respect to translation and rotation show that ∇^2 is the only simple differential operator that can come into play. That clear, the principle of correspondence with classical physics gives all the rest. One can hope that equally compelling considerations will fix the detailed mathematical form of the expression that we write down so far only symbolically,

$$\frac{\nabla^2 \psi}{(\delta^{(3)}\mathcal{G})^2}$$

However, a precise formulation of such considerations would seem to be out of reach until one has a knowledge of the tranformations of superspace comparable to one's knowledge of the transformations of 3-space. Hence the emphasis on the structure of superspace.

A problem of such depth can hardly be examined from too many points of view. Six more aspects of superspace seem worthy of mention. One can summarize them under the names "metric," "residual causality,"

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

"initial value," "conjugate momentum," "collapse," and "pregeometry." (1) There could hardly be a more helpful guide to the structure of superspace than the metric that obtains in it,

$$\left(\frac{1}{2g}\right)(g_{ik}g_{jl} + g_{il}g_{jk} - g_{ij}g_{kl})$$

Reference is made to [2] for the most illuminating discussion of this metric given to date. (2) This metric has a "light cone" associated with it. This light cone makes propagation proceed anisotropically in superspace. This anisotropy differs in character, however, from place to place according as ${}^{(3)}R$ is positive or negative. This anisotropy imposes a kind of "residual causality" upon superspace. Charles W. Misner has pointed out in a conversation that one cannot forget this residual causality when one says that the customary ideas of "before" and "after" lose their meaning at the scale of the Planck length. Perhaps one can say more: If the principle of causality has been of service in analyzing the structure of flat spacetime, it can hardly fail to help in studying the structure of superspace.

Tangent Vectors on Superspace and the Classical Initial Value Problem

(3) In the classical mechanics of a particle one is accustomed to specifying freely x_0 and $(dx/dt)_0$. These initial conditions determine the whole future history of the particle. What are the analogous freely disposable initial value data of classical geometrodynamics? One is tempted to say: conceive of a continuous one-parameter family of 3-geometries, specified for example in one coordinate patch by the 6 metric coefficients $g_{ik}(x, y, z; \lambda)$; and use this one-parameter family to define data analogous to x_0 and $(dx/dt)_0$ in the one-particle problem; thus, ${}^{(3)}\mathcal{G}_0$ stands for the class of metrics equivalent to $g_{ik}(x, y, z; 0)$; and $(d^{(3)}\mathcal{G}/d\lambda)_0$ is the "tangent vector in superspace" defined by $[\partial g_{ik}(x, y, z, \lambda)/\partial\lambda]_{\lambda=0}$, modulo the group of coordinate transformations.

It is easy for a mere coordinate shift to mock up the appearance of a change in geometry. Let the coordinates be shifted so that the point P, formerly characterized by the coordinates x^i, is now characterized by $x^i - \lambda\xi^i$, with the vector field ξ^i a continuous function of position. The metric g_{ik} is altered by this shift to

$$g_{ik} + \lambda(\xi_{i\,|\,k} + \xi_{k\,|\,i})$$

The derivative $(d^{(3)}\mathcal{G}/d\lambda)_0$ is $\xi_{i\,|\,k} + \xi_{k\,|\,i}$, modulo the group of coordinate transformations. But this quantity, by reason of its very origin, is obviously annullable by a coordinate transformation. Consequently there is in this

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

case no real change in the geometry. In other words, one cannot admit any otherwise reasonable-looking field of values for $(\partial g_{ik}/\partial\lambda)_0$ without running the risk of deception. One wants what has been called in [2] (GMD & IFS) a 3-geometry (of "acceptable" topology) and another "nearby" 3-geometry in order—one trusts (central hypothesis of the subject!)—to be able to determine the entire past and future of the space, and thereby a complete 4-geometry. But if one has been "deceived" in describing what is ostensibly a second and "nearby" 3-geometry, he may merely be repeating all over again the previously given 3-geometry. In that event he has only half the amount of initial value data needed to predict the dynamics. One can restate the situation in the following terms in the context of the 4-geometry (regarded temporarily as known!). A spacelike slice is made through the 4-geometry. That gives the one 3-geometry demanded as one of the essential ingredients of the initial value data. However, that one slice is not adequate to distinguish the given 4-geometry from any number of other, different 4-geometries that admit as slice the same 3-geometry. To complete the selection of the given 4-geometry from these alternative 4-geometries, erect vectors λn^α at each of the points of the 3-geometry, with n^α a continuous function of position. Their tips define a new hypersurface, the coordinates in which are connected continuously with the coordinates in the original hypersurface. Evaluate the metric coefficients $g_{ik}(x, y, z, \lambda)$ on this hypersurface. This hypersurface can be said to have been "pushed forward" with respect to the original hypersurface. But has it? Yes, if the normal component of λn^α nowhere vanishes. However, Hans Ohanian and Elliot Belasco have emphasized in unpublished remarks that it may happen that there are whole regions of the hypersurface where the normal component of λn^α vanishes. In that case one has not really pushed the hypersurface ahead at all in ${}^{(4)}\mathscr{G}$. In this event the second component of the initial value data, the derivatives $(\partial g_{ik}/\partial\lambda)_0$, will simply be inadequate for the purposes of the elliptic initial value equations [75]. This situation will be signaled by the fact that $(\partial g_{ik}/\partial\lambda)_0$ can be written in the form $\xi_{i\,|\,k} + \xi_{k\,|\,i}$. In other regions the 3-geometry *will* have been pushed forward in time. This situation will be signaled, except in special circumstances (time-symmetric initial value problem; change in g_{ik} proportional to λ^2 rather than λ; situation covered by appropriate care in the formulation), by the fact that $(\partial g_{ik}/\partial\lambda)_0$ (or $(\partial g_{ik}/\partial\lambda^2)_0$ in special cases like the time-symmetric initial value problem) is *not* representable in the form $\xi_{i\,|\,k} + \xi_{k\,|\,i}$.

It is natural to try to summarize the whole situation in the following form. *Give a point in superspace and give a "fully developed direction" at this point in superspace. Then (hypothesis!) this information is sufficient, together with Einstein's equations, uniquely to determine the entire 4-geometry.* Here the term "point in superspace" implies, as earlier, the demand that the 3-geometry in question have acceptable topology. The term "fully developed direction" implies that there is no point on the 3-geometry where the quantity

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$(\partial g_{ik}/\partial\lambda)_0$ (or if it vanishes, the quantity $(\partial g_{ik}/\partial\lambda^2)_0$) can be expressed in the form $\xi_{i\,|\,k} + \xi_{k\,|\,i}$. In brief, does superspace provide a new approach to the classical initial value problem? And in turn, does that initial value problem throw new light on the concept of "direction" in superspace?

(4) Superspace is built on the concept of 3-geometry; but dynamically conjugate to 3-geometry is the geometrodynamical momentum, with components π_{ij}. Out of these objects, with all the varied topologies that *they* can have, one can build a "conjugate superspace." What are *its* properties?

(5) No crisis stands out more insistently in all of physics than gravitational collapse. No topic connects so immediately the world of the very large and the very small. What insights can one gain from the concept of superspace into the cause and consequences of gravitational collapse?

(6) How far can one go in analyzing the properties of superspace without getting into the problems of "pregeometry"?

Level 3. Pregeometry

Weyl remarks [76] "... a more detailed scrutiny of a surface might disclose that, what we had considered an elementary piece, in reality has tiny handles attached to it which change the connectivity character of the piece, and that a microscope of even greater magnification would reveal ever new topological complications of this type, *ad infinitum*." Under such circumstances it would seem difficult to uphold the concept of dimensionality at the smallest distances. General arguments [77] emphasize the same point. Moreover, if electromagnetism and other fields have to do with the quantum-mechanical resonance of space between one topology and another, why should not the concept of metric itself be likewise a derived concept, going back for its foundation to topological or pretopological—and at any rate to pregeometric—ideas ("distance between A and B" being defined in the last analysis, for example, by "the ramification of the connections between A and B")? One cannot even mention these topics without recalling the universal sway of the quantum principle, and without stressing the "order of creation" as one thinks of it from physical evidence: Not first geometry and then the quantum principle, but first the quantum principle and then geometry!

It is enough to raise these issues, with all their depth, to see into what difficulties one can get with quantum geometrodynamics if one tries to think of it as an "ultimate" theory. However, physics has never depended for its progress on having an ultimate theory. There is no reason to think that the situation is different today. While one can raise ultimate issues of all kinds, there is no reason to believe that they all have to settled now! Nor that one *can* resolve them now! The subject presents an ever widening list of issues that have lively physical interest and lend themselves to well-known methods of analysis [78].

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PROBLEM 3. INITIAL CONDITIONS

The classical initial value problem has already been discussed. What can one say about the corresponding problem in quantum geometrodynamics? In other words, how much information must one give about $\psi({}^{(3)}\mathcal{G})$ on an appropriate submanifold of superspace in order to be able to predict this probability amplitude everywhere in superspace? And what is the character of this submanifold? In this connection one recalls that the "Einstein–Schroedinger wave equation" is of the second order. This second-order character raises a question of principle: In order to be able to calculate ψ everywhere, must one know on a hypersurface of superspace not only ψ but also its normal derivative? No, Leutwyler suggests in a most interesting paper [79]. He points out in the context of a simplified model that the natural features of superspace itself impose certain natural boundary conditions. They reduce the effective order of the equation from second to first.

Wider questions of principle are also posed by the very structure of quantum geometrodynamics. The arena of the dynamics is not space, but superspace. At first this development seems preposterous. How can one speak sensibly of any physical predictions when the outcome depends on what is taking place in unreachable regions of superspace? Nothing could seem more at variance with the spirit of science as dealing only with the knowable. However, a closer look shows that one has broken not at all with the traditional spirit of dynamics, but only with the details. In classical dynamics a clean distinction has always been maintained between (1) the equations of motion, which one can hope to know and understand, and (2) the origin of the initial conditions for those equations of motion—which is beyond one's power to investigate [80]. Quantum geometrodynamics maintains a similar cut between the knowable and the unknowable, but the cut comes in a new place [81]. Nothing seems to exclude the possibility ultimately to know (1) the detailed form of the Einstein–Schroedinger equation, and the concomitant structure of superspace; but as for (2) the source of the initial conditions on ψ, that would seem as far as ever beyond one's power ever to know. Happily, in neither classical dynamics nor quantum geometrodynamics does one have to know all initial conditions to make useful predictions! On the contrary, as Wigner has so often stressed [82], the role of physics is to predict the *correlations* between observations.

REFERENCES AND NOTES

1. L. Rosenfeld, *Annalen der Physik* **5**, 113 (1930) and *Z. Physik* **65**, 589 (1930) and *Annales de l'Institut Henri Poincaré* **2**, 25 (1932); P. G. Bergmann, *Phys. Rev.* **75**, 680 (1949); P. G. Bergmann and J. H. M. Brunings, *Rev. Mod. Phys.* **21**, 480 (1949); Bergmann, Penfield, Schiller, and Zatzkis, *Phys. Rev.* **78**, 329 (1950); P. A. M. Dirac,

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Can. J. Math. **2**, 129 (1950); F. A. E. Pirani and A. Schild, *Phys. Rev.* **79**, 986 (1950); P. Bergmann, *Helv. Phys. Acta*, Suppl. IV, 79 (1956), *Nuovo Cimento* **3**, 1177 (1956) and *Rev. Mod. Phys.* **29**, 352 (1957); C. W. Misner, *Rev. Mod. Phys.* **29**, 497 (1957); B. S. Dewitt, *Rev. Mod. Phys.* **29**, 377 (1957); P. A. M. Dirac, *Proc. Roy. Soc.* (*London*) **A246**, 326 and 333 (1958) and *Phys. Rev.* **114**, 924 (1959); B. S. DeWitt, "The Quantization of Geometry", a chapter in *Gravitation: An Introduction to Current Research* (L. Witten, ed), Wiley, New York, 1962; J. Schwinger, *Phys. Rev.* **130**, 1253 (1963) and **132**, (1317) (1963);

R. P. Feynman, mimeographed letter to V. F. Weisskopf dated 4 January to 11 February 1961; *Acta Physica Polonica* **24**, 697 (1963); *Lectures on Gravitation* (notes mimeographed by F. B. Morinigo and W. G. Wagner, California Institute of Technology, 1963); report in *Proceedings of the 1962 Warsaw Conference on the Theory of Gravitation* (PWN-Editions Scientifiques de Pologne, Warszawa, 1964); S. N. Gupta, report in *Recent Developments in General Relativity*, Pergamon, New York, 1962; S. Mandelstam, *Proc. Roy. Soc.* (*London*) **A270**, 346 (1962) and *Annals of Physics* **19**, 25 (1962); J. L. Anderson in *Proceedings of the 1962 Eastern Theoretical Conference* (M. E. Rose, ed.), Gordon and Breach, New York, 1963, p. 387; I. B. Khriplovich, "Gravitation and Finite Renormalization in Quantum Electrodynamics" (mimeographed report, Siberian Section Academy of Science, U.S.S.R., Novosibirsk, 1965); H. Leutwyler, *Phys. Rev.* **134**, B1155 (1964); B. S. DeWitt, "Dynamical Theory of Groups and Fields," in *Relativity Groups and Topology* (C. DeWitt and B. DeWitt, eds.), Gordon and Breach, New York, 1964; S. Weinberg, *Phys. Rev.* **135**, B1049 (1964), **138**, B988 (1965), and **140**, B516 (1965); M. A. Markov, *Progr. Theor. Phys., Yukawa Suppl.*, 1965, p. 85.

2. P. W. Higgs, *Phys. Rev. Letters* **1**, 373 (1958) and **3**, 66 (1959); R. Arnowitt, S. Deser, and C. W. Misner, a series of papers summarized in "The Dynamics of General Relativity" in *Gravitation: An Introduction to Current Research* (L. Witten, ed.), Wiley, New York, 1962; A. Peres, *Nuovo Cimento* **26**, 53 (1962); R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, *Phys. Rev.* **126**, 1864 (1962); cf. also the Princeton A. B. Senior Thesis of D. H. Sharp, May 1960 (unpublished); J. A. Wheeler, *Geometrodynamics*, Academic Press, New York, 1962, cited hereafter as GMD, and "Geometrodynamics and the Issue of the Final State," cited hereafter as GMD & IFS, a chapter in *Relativity, Groups and Topology* (C. DeWitt and B. DeWitt, eds.), Gordon and Breach, New York, 1964; B. DeWitt, *Phys. Rev.* **160**, 1113 (1967), **162**, 1195 (1967) and **162**, 1239 (1967), together cited hereafter as QTG.

3. J. A. Wheeler, GMD & IFS; B. DeWitt, QTG.

4. Michael D. Stern, *Investigations of the Topology of Superspace*, Princeton A. B. Senior Thesis, May 1967 (unpublished) and *Proc. Natl. Acad. Sci. U.S.A.* (submitted for publication).

5. R. Penrose, *An Analysis of the Structure of Spacetime*, Adams Prize Essay (mimeographed for limited distribution, Princeton University, Princeton, New Jersey, December 1966), gives a beautiful procedure to prescribe on the past light cone exactly enough geometrical information to determine out of Einstein's field equations the complete 4-geometry everywhere within the past light cone. The treatment is given only in the analytic case (in which case the difference between "inside" and "outside" the light cone does not make itself felt) but from general

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

considerations one must expect in the nonanalytic case that the data in question determine the 4-geometry only *within* the past light cone. When the light cone points into the future instead of the past, similar considerations of course apply, obtained by the interchange of the words "past" and "future" in what is said in the text. Problems arise with such formulations of the initial value problem "on the light cone" when the propagation proceeds far in a space of variable curvature. Then the light cone develops more than one sheet. Compare the several claps of thunder often heard from a single localized explosion!

6. For a discussion of the "observer" as a "collector of printout" see for example E. F. Taylor and J. A. Wheeler, *Spacetime Physics*, W. H. Freeman, San Francisco, 1966.

7. For the idea of a general spacelike hypersurface as the manifold on which the magnitudes of quantum field theory are to be measured, see especially S. Tomonaga, *Progr. Theor. Phys.* **1**, 34 (1946) and J. Schwinger, *Phys. Rev.* **74**, 1449 (1948).

8. J. A. Wheeler, GMD & IFS.

9. B. DeWitt, QTG.

10. This way of writing the conditions for constructive interference is symbolic only. In actuality the Hamilton–Jacobi function S depends not only upon the ${}^{(3)}\mathcal{G}$, but upon an infinity of parameters that distinguish one solution of the Hamilton–Jacobi equation from another. Thus, in a problem with one degree of freedom we write $S = S_0(x, E) + \delta(E)$ and in a problem with n degrees of freedom $S = S_0(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_n) + \delta(\alpha_1, \ldots, \alpha_n)$. In geometrodynamics there are two degrees of freedom per space point, the magnitudes associated with which may be designated by α and β. Thus the infinitude of freely disposable parameters may be indicated by two freely disposable functions, $\alpha(u, v, w)$ and $\beta(u, v, w)$. The ∞^3 points are given by the ∞^3 possible choices of u, v, and w. The u, v, w manifold may be, but is not required to be, the same as the manifold x, y, z of points in the 3-geometry ("alternative choices of parametrization of Hamilton–Jacobi function"). In any case we write S as a functional of α and β; thus, $S = S_0({}^{(3)}\mathcal{G};\ \alpha(u, v, w), \beta(u, v, w)) + \delta(\alpha(u, v, w), \beta(u, v, w))$. Then the conditions of constructive interference become statements about functional derivatives; thus,

$$\frac{\delta S}{\delta \alpha} = 0$$

and

$$\frac{\delta S}{\delta \beta} = 0$$

This is an explicit form of the symbolic Eq. (11) of the text. Other ways of writing the equations of constructive interference also exist.

11. Ulrich Gerlach, *Bull. Am. Phys. Soc.* for the Washington meeting of April, 1966, paper DE7, p. 340.

12. A. Peres, *Nuovo Cimento* **26**, 53 (1962). Here the unit of length is $(16\pi)^{1/2}L^* = (16\pi\hbar G/c^3)^{1/2}$

13. See [1] and [2].

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

14. R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, *Phys. Rev.* **126**, 1864 (1962).

15. The term "dimensionality" can be translated as the requirement of "imbeddability" of all the ${}^{(3)}\mathcal{G}$'s in a ${}^{(4)}\mathcal{G}$, a requirement that would seem the natural starting point for a derivation of the Einstein–Hamilton–Jacobi equation straight from first principles (see Problem 1).

16. C. W. Misner, *Rev. Mod. Phys.* **29**, 497 (1957); see also H. Leutwyler, *Phys. Rev.* **134**, B1155 (1964) and B. S. DeWitt, QTG.

17. M. Planck, *Sitzungsber. Preussische Akad. Wiss. Berlin, Math.-Phys. Klasse*, 1899, p. 440; J. A. Wheeler, GMD.

18. For a review of the subject of gravitational collapse, see for example B. K. Harrison, K. Thorne, M. Wakano, and J. A. Wheeler, *Gravitation Theory and Gravitational Collapse*, Univ. of Chicago Press, 1965; also A. G. Doroschkevich, Ya. B. Zel'dovich, and I. D. Novikov, *J. Exptl. Theor. Phys.* **49**, 170 (1965), English translation in *Soviet Physics JETP* **22**, 122 (1966) and Ya. B. Zel'dovich and I. D. Novikov, *Usp. Fiz. Nauk* **84**, 377 (1964) and **86**, 447 (1965), English translations in *Soviet Physics Usp.* **7**, 763 (1965) and **8**, 522 (1965).

19. For a presentation of the quantum electrodynamical calculation of the major part of the Lamb shift of hydrogen from this point of view, see T. A. Welton, *Phys. Rev.* **74**, 1157 (1948) and F. J. Dyson, *Advanced Quantum Mechanics*, Cornell Univ., Ithaca, 1954 (mimeographed), p. 54.

20. J. A. Wheeler, GMD.

21. J. A. Wheeler, GMD and GMD & IFS; B. S. DeWitt, QTG and "The Quantization of Geometry" in *Gravitation: An Introduction to Current Research* (L. Witten, ed.), Wiley, New York, 1962, p. 342 ff.

22. In GMD.

23. For an elementary discussion of the identity between the tide-producing component of the gravitational force and the Riemann curvature, see for example E. F. Taylor and J. A. Wheeler, *Spacetime Physics*, Freeman, San Francisco, 1966.

24. E. P. Wigner, "The Unreasonable Effectiveness of Mathematics in the Natural Sciences," in his book *Symmetries and Reflections*, Indiana Univ. Press, Bloomington, 1967; reprinted from *Comm. Pure App. Math.* **13**, No. 1 (February, 1960).

25. Reference 17.

26. A. S. Eddington, *Relativity Theory of Protons and Electrons*, Cambridge Univ. Press, 1936, and *Fundamental Theory*, Cambridge Univ. Press, 1946; also *Proc. Camb. Phil. Soc.* **27**, 15 (1931).

27. P. A. M. Dirac, *Nature* **139**, 323 (1937), *Proc. Roy. Soc.* (*London*) **A165**, 199 (1938).

28. P. Jordan, *Schwerkraft und Weltall*, Vieweg und Sohn, Braunschweig, 1955, and *Z. Physik* **157**, 112 (1959).

29. R. H. Dicke, *Science* **129**, 3349 (1959) and *The Theoretical Significance of Experimental Relativity*, Gordon and Breach, New York, 1964, p. 72.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

30. S. Hayakawa, *Progr. Theor. Phys.* **33**, 538 (1965) and *Progr. Theor. Phys. Suppl.* p. 532 (1965).

31. It is permissible to take at full force the argument of Eddington, Dirac, Jordan, Dicke, and Hayakawa, that a physical correlation exists between 10^{20}, 10^{40} and 10^{80}, without accepting the suggestion sometimes made in the same context, that the physical constants may "change with time." There is increasing observational evidence against such changes, and no incontrovertible evidence for them has ever been found. Among the relevant observations one can cite as examples R. H. Dicke, *Nature* **183**, 170 (1959) and *Nature* **192**, 440 (1961); also R. H. Dicke and P. J. E. Peebles, *J. Geophys. Res.* **67**, 10 and 4063 (1962) and *Phys. Rev.* **128**, 5 and 2006 (1962), showing no detectable change with time in the relative rates of selected processes of radioactive decay. To search for changes with time in the reciprocal fine structure constant, $\alpha^{-1} = \hbar c/e^2 = 137.03$, is simple in principle. One has only to compare the wavelength of the 21-cm line of hydrogen (red shifted because it was given out by a rapidly receding galaxy, far away and long ago) with the wavelength of a line in the optical spectrum (which has undergone the same red shift). The ratio R of the two wavelengths is α^{-1} multiplied by a known function of the atomic number of the source (an integer) and of the relevant quantum numbers (also integers):

$$R = \alpha^{-1} \text{ times function of integers}$$

The value of R for a source 1.4×10^9 light-years away (recession rate $\beta = v/c = 0.1$) should differ by several percent from the value of R for a laboratory source if any of the suggestions are correct that physical constants might change in proportion to the time (or some significant power of the time) measured from the start of the expansion of the universe. The writer is indebted to the kindness of Professor R. Minkowski of Berkeley for the following (July 28, 1967) summary of the observational situation: (1) Hopes have been dashed to observe the 21-cm line in the spectrum of galaxies anywhere near as far away as would correspond to a recession velocity of $\beta = 0.1$. (2) Observations have been made on 30 nearer objects by Dieter, Epstein, Lilley, and Roberts, *Astrophys. J.* **67**, 270 (1962) as supplemented by Roberts, *Astrophys. J.* **142**, 148 (1965). The red shift is the same for the 21-cm line and for the optical lines within the limits of error of the observations. It is difficult to evaluate the accuracy because individual motions amount to as much as 20–30% of the average recession velocity of $v = 1600$ km/sec ($\beta = 0.005$). It is probably safe to say that any change in α^{-1} must be less than a few percent to be compatible with the observations. However, a change anyway smaller than this limit would be expected on almost any of the varied theories of the change of α^{-1} with time, since the time lapse in this case is only one two-hundredth of the Hubble time. (3) Instead of comparing the wavelength of the 21-cm line with the wavelength of an optical transition, one can measure the fine structure separation of a related pair of lines in the optical spectrum itself. The fractional splitting $\Delta\lambda/\lambda$ should be independent of the red shift of the source if α^{-1} is constant. From the observations of R. Minkowski, *Astrophys. J.* **123**, 373 (1956), on Cygnus A ($v = 16830$ km/sec, $\beta = 0.056$: W. Baade and R. Minkowski, *Astrophys. J.* **119**, 206 (1954)), it is again probably safe to say that any change in the fine structure constant must be less than

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

a few percent. Bahcall, Sargent, and Schmidt, *Ap. J. Lett*, **149**, 11 (1967) give $|\Delta\alpha/\alpha| < 0.05$ for $z = 2$.

32. The quark, so useful in doing bookkeeping on the beautiful regularities of elementary particle physics (summarized, for example, in M. Gell-Mann and Y. Ne'eman, *The Eightfold Way*, Benjamin, New York, 1964 and F. Dyson, *Symmetry Groups in Nuclear and Particle Physics*, Benjamin, New York, 1966) has sometimes been taken much more seriously, as if it were an actual "primordial building block" of matter. That view may or may not be correct. If it is, and if one still continues to take quantum geometrodynamics as the only available indicator of what goes on at very small distances, then it would still seem reasonable to expect that one must have some perspective on what happens at 10^{-33} cm before one can find the rationale of quarks and particles. That there is no such thing as a quark in the literal sense is, however, a point of view accepted by many investigators, and stressed especially by Heisenberg and Dürr: W. Heisenberg, *Introduction to the Unified Field Theory of Elementary Particles*, Wiley, New York, 1966 and H. P. Dürr, "On the non-linear spinor theory of elementary particles," *Acta Physica Austriaca*, Suppl. III, 1966. Dürr has made the same point even more vividly (kind personal communication of June, 1967) by considering in effect what one would conclude out of the first several dozen atomic energy levels of an atom such as, for example, carbon or iron if one had (1) good measurements of the energies and transition probabilities, and (2) today's aptitude for searching for symmetries, but (3) not the slightest idea of the actual internal machinery of an atom. He shows how groups of high symmetry will make their appearance. His discussion leads one to ask whether the innocent investigator will not conclude that the atom is made out of quarks!

33. The strongest statement easily available against taking general relativity seriously at small distances appears to be that made by Robert Oppenheimer in his article "On Albert Einstein" (*New York Review*, March 17, 1966, pp. 4, 5): "He also worked with a very ambitious program, to combine the understanding of electricity and gravitation in such a way as to explain what he regarded as the semblance—the illusion—of discreteness, of particles in nature. I think that it was clear then, and believe it to be obviously clear today, that the things that this theory worked with were too meager, left out too much that was known to physicists but had not been known much in Einstein's student days. Thus it looked like a hopelessly limited and historically rather accidentally conditioned approach."

34. For further discussion of the rationale of changes in topology see Problem 2.

35. GMD.

36. To take it as self-evident that space is Euclidean in character at small distances became impossible after Riemann. His Göttingen inaugural lecture of June 10, 1854 pointed out that space can be highly rippled at submicroscopic distances and yet look smooth to all ordinary means of observation: "Uber die Hypothesen welche der Geometrie zugrunde liegen" in his *Gesammelte Mathematische Werke* (H. Weber, ed.), 2nd ed., reprinted by Dover, New York, 1953, also in a translation in *Nature* **8**, 14 (1873), by W. K. Clifford. Clifford himself went further in his lecture before the Cambridge Philosophical Society February 21, 1870, "On the Space-Theory of Matter," reprinted in his *Mathematical Papers* (R. Tucker, ed.),

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

London, 1882, also in his *Lectures and Essays* (L. Stephen and F. Pollock, eds.), Vol. 1, London, 1879. He proposed to consider a particle as made up of nothing but curved empty space, differing from the surrounding space precisely in this localized curvature—and perhaps also in its connectivity or local topology. In *Was ist Materie* (Springer, Berlin, 1924, esp. pp. 57, 58) Hermann Weyl again pointed out that space may be multiply connected in the small, and consequently: "The argument that the charge of the electron must be spread over a finite region, because otherwise it would possess infinite inertial mass, has thus lost its force. One cannot at all say, here is charge, but only, this closed surface encloses charge." He went on to comment that the enormous value of the ratio of electric to gravitation forces "seems to indicate that the total number of electrons in the universe is important for the constitution of the individual electron." Albert Einstein and Nathan Rosen, *Phys. Rev.* **48**, 73 (1935), proposed the concept of two nearly Euclidean spaces, connected here and there by thin bridges or tubes, through which electric lines of force thread, to give the appearance of charges of variegated signs in the "upper" space and corresponding charges of the opposite sign in the "lower" space. J. A. Wheeler, *Phys. Rev.* **97**, 511 (1955), reprinted in GMD, proposed instead the concept of a tube or handle or "wormhole" reaching between two different localities in one and the same Euclidean space. It is an automatic consequence of this picture that the universe should contain equal amounts of positive and negative electricity. It is another consequence, proved by Misner in 1957 straight from Maxwell's equations for empty space, that the charge, or flux of lines through the wormhole, must stay constant with time. The proof, C. W. Misner and J. A. Wheeler, *Annals of Physics* **2**, 525 (1957), reprinted in GMD, holds no matter how tortuously the lines of force may be twisted, no matter how wanting in symmetry the geometry of the wormhole may be, and no matter how violently the field and the geometry may subsequently change with time. Misner also showed here the beautiful ties that exist between the Maxwell theory in a multiply connected empty space and the mathematics of differential forms and homology groups. In his analysis the field and the geometry were assumed to evolve deterministically in time, in accordance with the classical equations of electrodynamics and geometrodynamics. Reasons out of fluctuation theory to consider "wormholes" a property, not of particles, but of all space, were first given by J. A. Wheeler in *Annals of Physics* **2**, 604 (1957), expanded in GMD.

37. No one has pointed out a more direct tie between the energy of vacuum fluctuations and macroscopic physics than H. B. G. Casimir, *Proc. Nederland Akad. Wetenschappen, Amsterdam* **60**, 793 (1948), who predicted a force between two parallel metal plates. No attempt is made here to cite the extensive literature that verifies the existence and the predicted magnitude of this force. The same kind of fluctuations that are verified by this force at macroscopic distances are also checked at distances $\sim 10^{-12}$ cm by the Lamb shift, the most impressive single development in quantum electrodynamics in the post-World War II period (Fig. 4 and [19]).

38. D. R. Brill and J. B. Hartle, *Phys. Rev.* **135**, B271 (1964).

39. For a historical survey that treats chemistry and atomic physics as the two parts of a single development, see for example: W. G. Palmer, *A History of the Concept of Valency to 1930*, Cambridge Univ. Press, 1965, and especially J. J. Lagowski, *The Chemical Bond*, Houghton Mifflin, Boston, 1966.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

40. On these regularities see for example the books cited in [32].

41. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* **108**, 1175 (1957).

42. The three issues listed here are taken up in more detail in the Appendix.

43. A. Einstein, in P. A. Schilpp, ed., *Albert Einstein: Philosopher Scientist* (Library of Living Philosophers, Evanston, Illinois, 1949, p. 81) remarks: "If one had the field-equation of the total field, one would be compelled to demand that the particles themselves would *everywhere* be describable as singularity-free solutions of the completed field-equations. Only then would the general theory of relativity be a *complete* theory."

44. E. Cartan, *Leçons sur la géométrie des espaces de Riemann*, Gauthier-Villars, Paris, 2nd ed., 1959, Chapter 8.

45. J. A. Wheeler, Chapter 4 on Cartan's geometrical interpretation of Einstein's field equations in *Gravitation and Relativity* (H. Y. Chiu and W. F. Hoffmann, eds.), Benjamin, New York, 1964.

46. Arguments that the propagator appropriate for any particle of spin two and mass zero necessarily has a leading term of the form

$$k^{-2}(g^{\mu\alpha}g^{\nu\beta} + g^{\mu\beta}g^{\nu\alpha} - g^{\alpha\beta}g^{\mu\nu})$$

are given by S. Weinberg, *Phys. Rev.* **138**, B988 (1965); also in expanded form in his contribution to S. Deser and K. W. Ford, eds., Brandeis Summer Institute in Theoretical Physics, 1964, Vol. 2, *Lectures on Particle and Field Theory*, Prentice-Hall, Englewood Cliffs, New Jersey, 1965.

47. See for example W. Pauli in "Die allgemeinen Prinzipien der Wellenmechanik" in *Handbuch der Physik* (Geiger and Scheel, eds.), Springer, Berlin, 1933, Vol. 24, part 1; reprinted in revised form in the new *Handbuch der Physik* (S. Flügge, ed.), Springer, Berlin, 1958, Vol. 5, part 1.

48. See the discussion of the problem of factor ordering in B. DeWitt, QTG; also the references to earlier discussions of this issue cited by him there.

49. For a determination of the wave equation from (1) the principle of Lorentz covariance and (2) a selection of one or another set of spin quantum numbers see E. P. Wigner, *Ann. of Math.* **40**, 149 (1939) and V. Bargmann and E. P. Wigner, *Proc. Natl. Acad. Sci. U.S.A.* **34**, 211 (1946), both reprinted in the collection *Symmetry Groups in Nuclear and Particle Physics* (F. J. Dyson, ed., Benjamin, New York, 1966.

50. For an analysis of the metric of superspace see B. DeWitt, QTG [2], and other work cited by DeWitt; see also S. Weinberg [46], for the propagator in the linear theory of gravitation in the de Donder gauge, which mysteriously has the same form as the metric in superspace (but indices 0, 1, 2, 3 as compared to 1, 2, 3).

51. For the proof that the topology of space cannot change within the context of classical geometrodynamics, see R. P. Geroch, *J. Math. Phys.* **8**, 782 (1967).

52. For a model of a closed universe with the topology S_3 put together out of 720 identical pieces each endowed with the Schwarzschild geometry ("lattice universe") see R. W. Lindquist and J. A. Wheeler, *Rev. Mod. Phys.* **29**, 432 (1957) and further treatment in GMD & IFS, pp. 370–379.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

53. A. Einstein, end of chapter dealing with Mach' principle in *The Meaning of Relativity*, Princeton Univ. Press, Princeton, New Jersey, 3rd ed., 1950.

54. See for example some of the solutions of Einstein's equations given by B. K. Harrison, *Phys. Rev.* **116**, 1285 (1959) and his fuller Princeton University Ph.D. thesis, *Exact Three-Variable Solutions of the Field Equations of General Relativity*, 1959 (unpublished).

55. Here it is assumed that the conjecture of H. Poincaré is correct, that every simply connected compact differentiable three-dimensional manifold is homeomorphic to the three-sphere. See C. D. Papakyriakopoulos, "The theory of differentiable manifolds since 1950," *Proc. Intern. Congr. Mathematicians, 1958* (Cambridge Univ. Press, 1960), pp. 433–440; also J. Milnor, *Topology from the Differentiable Viewpoint*, Univ. of Virginia Press, Charlottesville, 1965, and the bibliography cited by Milnor; also J. Derwent, "Handle decomposition of manifolds," *J. Math. Mech.* **15**, 329 (1966).

56. Particularly to be emphasized is the distinction between "asymptotically flat" as that concept is so often understood in the classical context of a 4-geometry, and the concept of flatness as it is applied to a 3-geometry in the context of Hamilton–Jacobi theory or quantum geometrodynamics. No example illustrates this distinction more clearly than the Schwarzschild geometry. There the rate of approach of the 4-geometry to flatness at infinity determines always a unique value for the mass of the center of attraction; but the analogous calculation for a spacelike 3-geometry slicing through the 4-geometry gives quite different values for the apparent mass, depending upon the choice of slice. Thus, in the 4-geometry

$$ds^2 = -(1 - 2m/r)dt^2 + (1 - 2m/r)^{-1}\, dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)$$

take at large distances the spacelike slice

$$t = t_0 + (8\alpha r)^{1/2}$$

so that

$$dt = \left(\frac{2\alpha}{r}\right)^{1/2} dr$$

On this slice one finds a 3-geometry in which the coefficient of dr^2, also at large distances, is

$$1 + \frac{2(m - \alpha)}{r}$$

The dependence of the "effective mass" $(m - \alpha)$ upon the choice of slice, through the parameter α, suggests some of the many dangers that seem to lurk in the concept of "asymptotic flatness" as applied to *three*-geometries.

57. In the Taub universe the effective mass-energy arises entirely from excitation of that mode of gravitational radiation which has the longest wavelength capable of fitting into this universe. For the metric of this model see A. Taub, *Ann. of Math.* **53**, 472 (1959) and C. W. Misner, *J. Math. Phys.* **4**, 924 (1963).

58. To say that a particle is "pictured in terms of space resonating from one topology to another" means more precisely that it is "pictured as a geometrodynam-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ical exciton—a state of excitation in which space resonates from one topology and geometry to another according to a probability amplitude function slightly different from, and orthogonal to, the probability amplitude function $\psi({}^{(3)}\mathcal{G})$ that describes the vacuum."

59. For a systematic development of "already unified field theory" see C. W. Misner and J. A. Wheeler, *Annals of Physics* **2**, 525 (1957) (reprinted in GMD) where also reference is made to the earlier work of G. Y. Rainich.

60. In "already unified field theory" the electromagnetic field tensor is expressed, in accordance with Einstein's field equations, in terms of the "Maxwell square root" of the Ricci curvature tensor and its dual

$$F_{\mu\nu} = (\text{“}\, R^{1/2}\,\text{”})_{\mu\nu} \cos\alpha + {}^*(\text{“}\, R^{1/2}\,\text{”})_{\mu\nu} \sin\alpha$$

The change of the "complexion" α of the electromagnetic field from place to place is fully determined by Maxwell's equations in places where there is a field. However, consider a spacelike initial value hypersurface. On this hypersurface consider two regions, I, II, endowed with field and separated by a region III free of field. Within each region individually the relative complexion is well determined, which is all that matters momentarily for the electrodynamics. However, the complexion of region II relative to region I can never be found from purely geometrical measurements limited to this initial spacelike hypersurface. Moreover, this relative complexion is all important for the dynamic development of the electromagnetic field at those later points in space-time that can be reached by disturbances both from I and from II. In this sense the initial value problem of already unified field theory does not lend itself to purely geometrical formulation. For more on this topic see the chapter by L. Witten in the book of which he is also the editor, *Gravitation: An Introduction to Current Research*, Wiley, New York, 1962.

61. The divergence condition is well known to follow from the invariance of the Hamilton–Jacobi function with respect to the gauge transformation, $A_i^{\text{new}} = A_i + \partial\lambda/\partial x^i$; thus,

$$0 = \delta S = \int \left(\frac{\delta S}{\delta A_i}\right) \delta A_i \, d^3x$$

$$= \int \left(\mathfrak{E}^i \frac{\partial \lambda}{\partial x^i}\right) d^3x$$

$$= -\int \mathfrak{E}^i_{,i} \lambda \, d^3x$$

The vanishing of this expression for arbitrary λ gives the desired relation. It should be emphasized that the quantity $\mathfrak{E}^i$ as employed here is not a contravariant vector, but $(g)^{1/2}$ times a contravariant vector ("vector density"). Were the contravariant vector itself employed, the divergence relation would have to be expressed in terms of covariant derivatives rather than ordinary derivatives, complicating the derivation in the text.

62. Here the symbol $(\mathfrak{E} \times B)_j$ stands for the covariant vector density $\mathfrak{E}^i B_{ij}$.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

63. As Hermann Weyl emphasized long ago, *Math. Z.* **23**, 271 (1925), "In den geometrischen und physikalischen Anwendungen zeigte sich stets, dass eine Grössart nicht allein durch Angabe der Tensorstufe, sondern durch Symmetrie bedingungen charakterisiert ist." In other words, every physical quantity is represented by an irreducible tensorial quantity; that is to say, by what S. S. Chern terms "a geometrical object." Weyl conceived of these geometrical entities as local. However, it is a natural extension of his line of thought to speak of a *functional* S or ψ that depends *globally* upon a 3-geometry, and upon a 2-form imbedded in that 3-geometry.

64. GMD [2], p. 88.

65. John Milnor, "A survey of cobordism theory," *L'enseignement mathématique* **8**, 16 (1962); "Spin structures on manifolds," *ibid.* **9**, 198 (1963); "On the Stiefel–Whitney numbers of complex manifolds and of spin manifolds," *Topology* **3**, 223 (1965); "Remarks concerning spin manifolds" in S. S. Cairns, ed., *Differential and Combinatorial Topology*, Princeton Univ. Press, Princeton, New Jersey, 1965, p. 55; Andre Lichnerowicz, *Compt. Rend. Acad. Sci. Paris* **252**, 3742 (1961), **253**, 940 (1961), and **253**, 983 (1961), and summary of these results in the third part of the chapter by Lichnerowicz, "Propagateurs, Commutateurs et Anticommutateurs en Relativité Générale," in C. and B. DeWitt, eds., *Relativity, Groups and Topology*, Gordon and Breach, New York, 1964; D. W. Anderson, E. H. Brown, Jr., and F. P. Peterson, "Spin cobordism," *Bull. Am. Math. Soc.* **72**, 256 (1966); "SU-cobordism, KO-characteristic numbers and the Kervaire invariant," *Ann. of Math.* **83**, 54 (1966); W. C. Hsiang and B. J. Sanderson, "Twist-spinning spheres in spheres," *Illinois J. Math.* **9**, 651 (1965). Appreciation is expressed to John Milnor, Roger Penrose, and Robert Geroch for discussions clarifying the concept of spin structure.

66. Take a belt. Stretch it out flat and taut. Keeping the left-hand end A fixed in the left hand, twist the right-hand end B through 720°. Maintaining A and B all the time parallel to their present orientations, move B in a complete circle about A (releasing for an instant one's hold on B). The belt straightens out. Not so when there is only a 360° twist in it. The belt is relevant to the cube, the room, and the eight elastic strings. Before the cube is rotated at all, it can be pulled out through a window to some distance from the room. The eight elastic strings then take on the configuration of the belt. The distinction between 360° and 720° rotation for the belt applies equally to the "pseudo-belt" made up of the eight strings. Another way of seeing that a 720° rotation restores the orientation entanglement relation between the cube and its surroundings (picture of one cone rolling on another) is presented by R. Penrose and W. Rindler in a preprint of an appendix to a book that they have in preparation. Appreciation is expressed to Professor Penrose for the privilege of seeing this preprint.

67. The possibility has been suggested that one may eventually be able to detect what is here called the "orientation entanglement relation" between an object and its surroundings by measuring the contact potential between one metallic object (subject to rotation) and another (held fixed): Y. Aharonov and L. Susskind, *Phys. Rev.* **158**, 1237 (1967).

68. Cf. section on orientability, Problem 2, Level 1.

69. It is not new to abstractify geometry. Einstein's curved space-time was in the beginning nothing if it was not a home for geodesics. How else, one asked,

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

could he predict a planetary motion. Later Einstein, Grommer, Infeld, and Hoffman threw out the geodesics. The field equations themselves, they showed, predict the evolution of geometry with time, and hence the motion of concentrations of mass-energy.

70. For a history of the concept of spin, see W. Pauli's 1945 Nobel prize lecture, *Exclusion Principle and Quantum Mechanics*, Editions Grisson, Neuchatel, 1947, also the relevant discussion in M. Fierz and V. F. Weisskopf, *Theoretical Physics in the Twentieth Century: A Memorial Volume to Wolfgang Pauli*, Interscience, New York, 1960.

71. See for example C. Kittel, *Introduction to Solid-State Physics*, 3rd ed., Wiley, New York, 1966.

72. The torus can be converted into a single sheet by two cuts, and can then be conceived as laid out on the complex plane, with one corner at the origin. One adjacent corner can be identified arbitrarily with the number $1 + 0i$, by appropriate choice of scale ("conformal transformation"). The location of the other adjacent corner $\tau = \tau_1 + i\tau_2$ is then completely determined. Also completely determined is the concomitant 2-geometry, modulo the group of conformal transformations. The quantity τ can be identified with the complex parameter mentioned in the text.

73. For a survey, with many references to the literature, see H. E. Rauch, "A transcendental view of the space of algebraic Riemann surfaces," *Bull. Am. Math. Soc.* **71**, 1 (1965). Appreciation is expressed to Professor Leon Ehrenpreis for elucidation of the subject and for this and the following reference.

74. Lipman Bers, *On the Moduli of Riemann Surfaces*; lectures at the Forschunginstitut für Mathematik, Eidgenössische Technische Hochschule, Zürich, 1964. Notes by L. M. and R. J. Sibner (mimeographed).

75. For the initial value equations of classical geometrodynamics see G. Darmois, *Les équations de la gravitation einsteinienne*, Gauthier-Villars, Paris, 1927; K. Stellmacher, *Math. Ann.* **115**, 136 (1937); A. Lichnerowicz, *J. math. pure appl.* **23**, 37 (1944); *Helv. Phys. Acta Suppl.* **4**, 176 (1956); *Théories relativistes de la gravitation et de l'électromagnétisme*, Masson, Paris, 1955; Yvonne Fourés-Bruhat, *Acta Math.* **88**, 141 (1952); *J. Rational Mech. Anal.* **5**, 951 (1956); and the chapter by Y. Fourés (now Y. Choquet) in Louis Witten, ed., *Gravitation: An Introduction to Current Research*, Wiley, New York, 1962.

76. H. Weyl, *Philosophy of Mathematics and Natural Science* (original German in 1927; translation by O. Helmer), Princeton Univ. Press, Princeton, New Jersey, 1949, p. 91.

77. GMD & IFS, pp. 495–499.

78. An extensive list of problems open for further investigation is to be found in GMD & IFS.

79. H. Leutwyler in *Battelle Rencontres: 1967 Lectures in Mathematics and Physics*, this volume p. 309.

80. One is reminded in this connection of the statement of William James over a half a century ago, that "Actualities seem to float in a wider sea of possibilities

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

from out of which they were chosen; and *somewhere*, indeterminism says, such possibilities exist, and form a part of truth." Appreciation is expressed to Paul Van de Water for this quotation.

81. E. P. Wigner, *Symmetries and Reflections*, Indiana Univ. Press, Bloomington, Indiana, 1967; see J. M. Jauch, E. P. Wigner, and M. M. Yanase, *Nuovo Cimento* **48**, 144 (1967), also B. DeWitt, QTG; also H. Everett III, *Rev. Mod. Phys.* **29**, 454 (1957); J. A. Wheeler, *Rev. Mod. Phys.* **29**, 463 (1957) and GMD, p. 75.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Reprinted from THE PHYSICAL REVIEW, Vol. 160, No. 5, 1113–1148, 25 August 1967
Printed in U. S. A.

Quantum Theory of Gravity. I. The Canonical Theory*

BRYCE S. DEWITT
Institute for Advanced Study, Princeton, New Jersey
and
Department of Physics, University of North Carolina, Chapel Hill, North Carolina†
(Received 25 July 1966; revised manuscript received 9 January 1967)

Following an historical introduction, the conventional canonical formulation of general relativity theory is presented. The canonical Lagrangian is expressed in terms of the extrinsic and intrinsic curvatures of the hypersurface x^0 = constant, and its relation to the asymptotic field energy in an infinite world is noted. The distinction between finite and infinite worlds is emphasized. In the quantum theory the primary and secondary constraints become conditions on the state vector, and in the case of finite worlds these conditions alone govern the dynamics. A resolution of the factor-ordering problem is proposed, and the consistency of the constraints is demonstrated. A 6-dimensional hyperbolic Riemannian manifold is introduced which takes for its metric the coefficient of the momenta in the Hamiltonian constraint. The geodesic incompletability of this manifold, owing to the existence of a frontier of infinite curvature, is demonstrated. The possibility is explored of relating this manifold to an infinite-dimensional manifold of 3-geometries, and of relating the structure of the latter manifold in turn to the dynamical behavior of space-time. The problem is approached through the WKB approximation and Hamilton-Jacobi theory. Einstein's equations are revealed as geodesic equations in the manifold of 3-geometries, modified by the presence of a "force term." The classical phenomenon of gravitational collapse shows that the force term is not powerful enough to prevent the trajectory of space-time from running into the frontier. The as-yet unresolved problem of determining when the collapse phenomenon represents a real barrier to the quantum-state functional is briefly discussed, and a boundary condition at the barrier is proposed. The state functional of a finite world can depend only on the 3-geometry of the hypersurface x^0 = constant. The label x^0 itself is irrelevant, and "time" must be determined intrinsically. A natural definition for the inner product of two such state functionals is introduced which, however, encounters difficulties with negative probabilities owing to the barrier boundary condition. In order to resolve these difficulties, a simplified model, the quantized Friedmann universe, is studied in detail. In order to obtain nonstatic wave functions which resemble a universe evolving, it is necessary to introduce a clock. In order that the combined wave functions of universe-*cum*-clock be normalizable, it turns out that the periods of universe and clock must be commensurable. Wave packets exhibiting quasiclassical behavior are constructed, and attention is called to the phenomenological character of "time." The inner-product definition is rescued from its negative-probability difficulties by making use of the fact that probability flows in a closed finite circuit in configuration space. The article ends with some speculations on the uniqueness of the state functional of the actual universe. It is suggested that a viewpoint due to Everett should be adopted in its interpretation.

1. INTRODUCTION

ALMOST as soon as quantum field theory was invented by Heisenberg, Pauli, Fock, Dirac, and Jordan, attempts were made to apply it to fields other than the electromagnetic field which had given it—and indeed quantum mechanics itself—birth. In 1930 Rosenfeld[1] applied it to the gravitational field which, at the time, was still regarded as the *other* great entity of Nature. Rosenfeld was the first to note some of the special technical difficulties involved in quantizing gravity and made some early attempts to develop general methods for handling them. As an application of his methods he computed the gravitational self-energy of a photon in lowest order of perturbation theory. He obtained a quadratically divergent result, confirming that the divergence malady of field theory, which had already been discovered in connection with the electron's electromagnetic self-energy, was widespread and deep seated. It is tempting, and perhaps no longer premature, to read into Rosenfeld's result a forecast that quantum gravidynamics was destined, from the very beginning, to be inextricably linked with the difficult issues lying at the theoretical foundations of particle physics.

During physics's great boom of the thirties the difficult issues of field theory were inevitably often bypassed. Moreover, it was recognized early that as far as the gravitational field is concerned its quanta (assuming they exist) can produce no *observable* effects until energies of the order of 10^{28} eV are reached, this fantastic energy corresponding to the so-called "Planck length" $(hG/c^3)^{1/2} \approx 10^{-33}$ cm, where G is the gravitation constant. Hence, after Rosenfeld's initial studies years passed before anything essentially new was done in quantum gravidynamics, and even today interest in this area of research is confined to a very small group of workers.

In 1950 the author[2] reperformed Rosenfeld's self-energy calculation in a manifestly Lorentz-covariant and gauge-invariant manner. This work was stimulated by the then new "renormalization" methods, which

* This research was supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR-153-64.
† Permanent address.
[1] L. Rosenfeld, Ann. Physik 5, 113 (1930); Z. Physik 65, 589 (1930).
[2] B. S. DeWitt, Ph.D. thesis, Harvard University, 1950 (unpublished).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education,Bantalab Jammu

had been developed by Tomonaga, Schwinger, and Feynman, and had as its aim a demonstration that Rosenfeld's result implies merely a renormalization of charge rather than a nonvanishing photon mass. An unanticipated source of potential difficulty arose in this calculation from the fact that not one but *two* gauge groups are simultaneously present (the group associated with gravity in addition to the familiar electromagnetic group) and that these groups are not combined in the form of a direct product but rather in the form of a *semidirect product* based on the automorphisms of the electromagnetic gauge group under general coordinate transformations. This means that if a fixed choice of gauge is to be maintained, every coordinate transformation must be accompanied by an electromagnetic gauge transformation. The calculation was pushed, however, again only to the lowest order of perturbation theory; in this order, which involves only single closed Feynman loops, the ensuing complications are easily dealt with.

At about the same time investigations of a more ambitious kind were undertaken by Bergmann.[3] Although the renormalization philosophy had proved a resounding success in quantum electrodynamics it was still under critical attack because the methods then (and frequently even now) in use involved the explicit manipulation of divergent quantities. Similar (although more elementary) difficulties also persisted in classical particle theories with one important exception, namely, the theory of the interaction of point masses with gravity. In 1938, Einstein, Infeld, and Hoffmann[4] had shown that the laws of motion of such particles follow from the gravitational field equations alone, without divergent quantities ever appearing or such concepts as self-mass intervening at any time. Moreover, this result had been subsequently extended to include electrically charged particles, and gave promise of being applicable to spinning particles as well. The gravitational field thus appeared as a kind of classical regulator, and Bergmann reasoned that the same might be true in the quantum theory. Since the fields are basic, in the Einstein-Infeld-Hoffmann view, and the particles are merely singularities in the fields, Bergmann's first task was to quantize the gravitational field. It was to be hoped that commutation relations for particle position and momentum would then follow as corollaries.

The obstacles which Bergmann faced were enormous. First of all, since the laws of particle motion depend crucially on the nonlinear properties of the Einstein field equations, it was necessary to quantize the full nonlinear gravitational field. Secondly, it was necessary to find some way of defining particle position and momentum in terms of field variables alone. Thirdly, it would eventually be necessary to include spin, so that quantized particles obeying a Dirac-like equation could be described. Fourthly, it would be necessary to extract Fermi statistics (for the particles) out of the Bose statistics obeyed by the gravitational field. Finally, it would be necessary ultimately to remove the asymmetry between particle and field inherent in the Einstein-Infeld-Hoffmann approach, so as to be able, as in quantum electrodynamics, to account for pair production and vacuum polarization. It is not surprising that Bergmann's goal today remains as elusive as ever.

To achieve this goal Bergmann set out upon the classical canonical road in search of a Hamiltonian. Despite the fact that canonical methods, by singling out the time for special treatment, run counter to the spirit of any relativistic theory—above all, such a completely covariant theory as general relativity—such a procedure seemed a good one for several reasons. Firstly, no other method was then known. Secondly, canonical methods afford quick insights into certain aspects of any theory. Thirdly, it seemed that standard perturbation methods would become available for certain types of calculations.

However, Bergmann immediately ran into major difficulties (some of which had already been foreseen by Rosenfeld) in the first stages of his program. These are referred to as "the problem of *constraints*," and are manifested in the following ways: Some of the field variables possess no conjugate momenta; the momenta conjugate to the remaining field variables are not all dynamically independent; the field equations themselves are not linearly independent, and some of them involve no second time derivatives, thus complicating the Cauchy problem. These difficulties are all related and arise from the existence of the general coordinate-transformation group as an invariance group for the theory.

Similar difficulties had already been encountered with the electromagnetic field and methods for handling them were well known. The same methods, however, proved to be much more difficult to apply in the case the gravitational field. An obstacle is created, for example, by the fact that not all of the relations between the momenta (i.e., the constraints) are linear. Moreover, because the invariance group of gravity is non-Abelian (in contrast to the gauge group of electrodynamics) tedious calculations must be performed to check that the commutators of the various constraints lead to no inconsistencies.

Bergmann and his co-workers performed much valuable ground work in formulating the difficulties in a precise way and in partially resolving them. In the meantime additional help came from an unexpected quarter. In 1950 Dirac[5] published the outline of a general Hamiltonian theory which is in principle applicable to any system describable by an action functional. Dirac's methods were quickly seized upon by Pirani and Schild[6]

[3] A bibliography of Bergmann's early work will be found in P. G. Bergmann, Helv. Phys. Acta Suppl. 4, 79 (1956).
[4] A. Einstein, L. Infeld, and B. Hoffmann, Ann. Math. 39, 65 (1938).
[5] P. A. M. Dirac, Can. J. Math. 2, 129 (1950).
[6] F. A. E. Pirani and A. Schild, Phys. Rev. 79, 986 (1950).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

for application to the gravitational field. Unfortunately, these authors chose to develop the theory within the framework of a "parameter formalism," in the hope, which eventually proved to be misplaced, of retaining a manifest covariance which Dirac's methods would otherwise destroy. The complexity of the resulting algebra prevented them from computing all of the constraints.

The theory remained in this incomplete state for several years. It was not until impetus was provided by the first international relativity conference in Bern in 1955 (Jubilee of Relativity Theory) and the second one in Chapel Hill in 1957 that things began to move again. A small step forward was made by the author,[7] who showed, using the Pirani-Schild formalism, that the four so-called "primary" constraints could, by a phase transformation, be changed into pure momenta. This meant that the state functional for gravity must be independent of the $g_{0\mu}$ components of the metric tensor ($\mu=0, 1, 2, 3$). Shortly afterward Higgs[8] showed that three of the so-called "secondary" or "dynamical" constraints are the generators of infinitesimal transformations of the three "spatial" coordinates x^1, x^2, x^3. The implication of this was that the state functional must be independent of the coordinates chosen in the spacelike cross sections $x^0=$ constant and hence cannot be taken to be an *arbitrary* functional of the metric components g_{ij} ($i, j=1, 2, 3$). Developments thereafter came rapidly. Dirac himself had by this time begun to apply his methods to the gravitational field.[9] As a result of simplifications and clarifications which he introduced, it became easy to show that the fourth dynamical constraint is consistent with the others, and the formal theory achieved for the first time a state of technical completion. It was then possible to begin asking "What does it all mean?"

On the classical side, the problems of physical interpretation were soon resolved by the work of Arnowitt, Deser, and Misner,[10] who showed how to use the canonical theory to provide a rigorous characterization of gravitational radiation and "energy." In the quantum domain, however, the interpretation of the formalism remained puzzling and obscure for several years, because one did not know the right questions to ask. It is only recently that the relevant issues have begun to come into focus, largely as a result of the patient researches of Wheeler,[11] whose ideas have proved a great source of stimulation to many workers, including the author.

[7] Reported at a meeting at Stevens Institute of Technology in January, 1958 (unpublished).

[8] P. W. Higgs, Phys. Rev. Letters 1, 373 (1958); 3, 66 (1959).

[9] P. A. M. Dirac, Proc. Roy. Soc. (London) A246, 326 (1958); A246, 333 (1958); Phys. Rev. 114, 924 (1959).

[10] R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962).

[11] The work of Wheeler and his associates is well described in J. A. Wheeler, *Relativity Groups and Topology, 1963 Les Houches Lectures* (Gordon and Breach Science Publishers, Inc., New York, 1964). This reference contains a large bibliography of additional papers on quantization, collapse, and many other related topics.

The present paper is the direct outcome of conversations with Wheeler,[12] during which one fundamental question in particular kept recurring: *What is the structure of the domain manifold for the quantum-mechanical state functional?* The attempt to answer this question has required a more far-reaching analysis of the technical structure of the canonical theory than can be found in the previous literature. The results of this analysis are here presented and used to develop an interpretative framework which, although tentative, is perhaps capable of serving in a variety of contexts.

Attention is mainly confined to the case of closed finite worlds, firstly because the issues which finite worlds raise are more critical and bizarre, and secondly because the case of infinite worlds is better handled within the framework of the so-called manifetlsy covariant theory which will be treated in two subsequent papers of this series. The latter theory, which has also achieved a state of technical completion following the pioneering work of Feynman,[13] differs utterly in its structure from the canonical theory, and so far no one has established a rigorous mathematical link between the two. At the present time the two theories play complementary roles, the canonical theory describing the quantum behavior of 3-space regarded as a time-varying geometrical object, and the covariant theory describing the behavior of real and virtual gravitons propagating in this object.

Section 2 of the present paper begins with the derivation of the canonical Lagrangian. Its structure in terms of the extrinsic and intrinsic curvatures of the hypersurface $x^0=$ constant is displayed, and attention is called to its relation to the total field energy in an asymptotically flat world. Section 3 is devoted to the primary and secondary constraints of the theory and to the independent question of coordinate conditions. Quantization is introduced in Sec. 4. Here the puzzling question of the role of a vanishing Hamiltonian is resolved by emphasizing the distinction between finite and infinite worlds. Asymptotic energy is an indispensable concept in an infinite world, and the Hamiltonian must be chosen accordingly. In a finite world there is no asymptotic energy, and an intrinsic description of the dynamics must be found, based on the constraints alone. The consistency of the constraints is demonstrated by straightforward computation of their commutators. The factor-ordering problem is disposed of by formal arguments which in effect assert that field variables taken at the same space-time point should be regarded as freely commutable. The "χ constraints" are shown to be the generators of 3-dimensional coordinate transformations.

In Sec. 5 the metric representation is introduced. The

[12] Any errors or wrong conjectures it contains are the author's own.

[13] R. P. Feynman, Mimeographed letter to V. F. Weisskopf dated January 4 to February 11, 1961 (unpublished); Acta Phys. Polon. 24, 697 (1963).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

distinction between finite and infinite worlds is again noted, and it is emphasized that the state functional in the former case depends only on the 3-geometry of the hypersurface $x^0 =$ constant and not on the label x^0 itself. The concept of a manifold $\mathfrak{M}$ of 3-geometries is introduced, and the role played by the Hamiltonian constraint in determining its geometrical structure is suggested. The coefficient of the momenta in the Hamiltonian constraint may be regarded as a metric of a 6-dimensional hyperbolic Riemannian manifold M. The structure of this manifold is studied in detail, and its geodesic incompletability, owing to the existence of a frontier of infinite curvature, is noted. The possibility of relating M to the question of "intrinsic time" for the state functional is discussed, and a natural definition for the inner product of two state functionals is proposed.

In Sec. 6 a natural metric based on M is assigned to the infinite-dimensional manifold $\mathfrak{M}$, and some of the properties of geodesics in $\mathfrak{M}$ are examined. An attempt is then made to indicate the extent to which the dynamical properties of the quantized gravitational field are determined by the structure of $\mathfrak{M}$. The attempt is heuristic and far from complete, and much work remains to be done. The problem is approached through the WKB approximation and Hamilton-Jacobi theory. Einstein's equations are revealed as geodesic equations in $\mathfrak{M}$, modified by the presence of a "force term." The classical phenomenon of gravitational collapse shows that the force term is not powerful enough to prevent the trajectory of 3-space from striking the frontier of $\mathfrak{M}$. The problem of determining when the collapse phenomenon represents a real barrier to the quantum state functional is briefly discussed, and a boundary condition (vanishing state functional) at the barrier is proposed.

The barrier boundary condition raises difficulties with the definition of probability. In order to study these difficulties it is useful to test the theory on a simplified model. In Sec. 7 the quantized Friedmann universe is studied in detail, and its static wave functions in the WKB approximation are obtained. In order to obtain nonstatic wave functions which resemble a dynamical universe evolving it is necessary to introduce a clock. The combined wave functions of universe-*cum*-clock are studied, and it is pointed out that normalizability of the wave functions requires precise commensurability between the periods of universe and clock.

Wave packets exhibiting quasiclassical behavior are constructed in Sec. 8, in three different representations. Two of these make use of proper times defined by the clock and the universe respectively; the third treats universe and clock symmetrically through their mutual correlations. Attention is called to the deficiencies of the first two representations arising from the fact that, in a covariant theory, time is only a phenomenological concept. In the third representation probability flows in a closed finite circuit in configuration space, and wave packets do *not* ultimately spread in time. Use is made of this fact in Sec. 9 to show how the inner-product definition can be rescued from the negative probability difficulties arising from the barrier boundary condition $\Psi = 0$ at $R = 0$ ($R =$ radius of universe). It is also shown that the conventional Cauchy data for the wave function suffice to determine the quantum state completely.

Section 10 is devoted to speculations on the general theory. An interpretation of quantum mechanics due to Everett (see Ref. 52) is described and proposed for dealing with the concept of "a wave function for the universe." Such an interpretation is essential if the wave function is unique. Evidence is presented that the Hamiltonian constraint may indeed have only one solution. The problem of time-reversal invariance and entropy is briefly discussed. Two technical appendices follow at the end of the article.

Attention is called to the following points of notation: Latin indices range over the values 1, 2, 3 and Greek indices over the values 0, 1, 2, 3. Differentiation is denoted by a comma. The coordinates x^0 and x^i are assumed to be timelike and spacelike, respectively, and the geometry of space-time is assumed to be such that the hypersurfaces $x^0 =$ constant are capable of carrying a complete set of Cauchy data. So-called "absolute units" in which $\hbar = c = 16\pi G = 1$ (G being the gravitation constant) are used throughout, as is also the signature $-+++$ for the space-time metric $g_{\mu\nu}$. The Riemann and Ricci tensors, and the curvature scalar, are taken in the respective forms

$$R_{\mu\nu\sigma}{}^{\tau} \equiv \Gamma_{\nu\sigma}{}^{\tau}{}_{,\mu} - \Gamma_{\mu\sigma}{}^{\tau}{}_{,\nu} + \Gamma_{\nu\sigma}{}^{\rho}\Gamma_{\mu\rho}{}^{\tau} - \Gamma_{\mu\sigma}{}^{\rho}\Gamma_{\nu\rho}{}^{\tau}, \quad (1.1)$$

$$R_{\mu\nu} \equiv R_{\sigma\mu\nu}{}^{\sigma}, \quad (1.2)$$

$${}^{(4)}R \equiv R_{\mu}{}^{\mu} \equiv g^{\mu\nu}R_{\mu\nu}, \quad (1.3)$$

where

$$\Gamma_{\mu\nu}{}^{\sigma} \equiv \tfrac{1}{2}g^{\sigma\tau}(g_{\mu\tau,\nu} + g_{\nu\tau,\mu} - g_{\mu\nu,\tau}), \quad g_{\mu\sigma}g^{\sigma\nu} = \delta_{\mu}{}^{\nu}. \quad (1.4)$$

The corresponding tensors in the spacelike cross sections $x^0 =$ constant are distinguished by means of a prefixed superscript (3). These conventions have the property that ${}^{(4)}R$ is non-negative in a space-time containing normal matter and satisfying Einstein's equations, and that ${}^{(3)}R$ is positive in a 3-space of positive curvature.

2. EXTRINSIC AND INTRINSIC CURVATURE. CLASSIC FORM OF THE LAGRANGIAN

The canonical theory begins with the following decomposition of the metric tensor:

$$(g_{\mu\nu}) = \begin{pmatrix} -\alpha^2 + \beta_k\beta^k & \beta_j \\ \beta_i & \gamma_{ij} \end{pmatrix}, \quad (g^{\mu\nu}) = \begin{pmatrix} -\alpha^{-2} & \alpha^{-2}\beta^j \\ \alpha^{-2}\beta^i & \gamma^{ij} - \alpha^{-2}\beta^i\beta^j \end{pmatrix}, \quad (2.1)$$

$$\gamma_{ik}\gamma^{kj} = \delta_i{}^j, \quad \beta^i = \gamma^{ij}\beta_j. \quad (2.2)$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

When the conventional Einstein Lagrangian density is reexpressed in terms of the new variables, it is found, after some calculation, to take the form

$$\mathfrak{L} \equiv g^{1/2}\,{}^{(4)}R = \alpha\gamma^{1/2}(K_{ij}K^{ij} - K^2 + {}^{(3)}R) - 2(\gamma^{1/2}K)_{,0} + 2(\gamma^{1/2}K\beta^i - \gamma^{1/2}\gamma^{ij}\alpha_{,j})_{,i} \quad (2.3)$$

where

$$g \equiv -\det(g_{\mu\nu}) = \alpha^2\gamma, \qquad \gamma \equiv \det(\gamma_{ij}), \quad (2.4)$$

$$K_{ij} \equiv \tfrac{1}{2}\alpha^{-1}(\beta_{i,j} + \beta_{j,i} - \gamma_{ij,0}), \quad K^{ij} \equiv \gamma^{ik}\gamma^{jl}K_{kl}, \quad K \equiv \gamma^{ij}K_{ij}, \quad (2.5)$$

the dots denoting covariant differentiation based on the 3-metric γ_{ij}.

The quantity K_{ij}, which transforms as a symmetric tensor under spatial coordinate transformations, is known as the *second fundamental form*. It describes the curvature of the hypersurface $x^0 =$ constant as viewed from the 4-dimensional space-time in which it is embedded. It is therefore also frequently called the *extrinsic curvature tensor* of the hypersurface, as opposed to the *intrinsic curvature tensor* ${}^{(3)}R_{ij}$, which depends only on γ_{ij} *in* the hypersurface. In a flat space-time ${}^{(3)}R_{ij}$ is completely determined by K_{ij}, but in a manifold of arbitrary curvature there need be no relationship between the two. The contracted forms ${}^{(3)}R$ and $K_{ij}K^{ij} - K^2$ will be referred to as the intrinsic and extrinsic curvatures, respectively.

The last two terms of Eq. (2.3), being total derivatives, are dynamically irrelevant and may be dropped. The Lagrangian then becomes

$$L \equiv \int \alpha\gamma^{1/2}(K_{ij}K^{ij} - K^2 + {}^{(3)}R)\,d^3x, \quad (2.6)$$

which has the classic form "kinetic energy minus potential energy," with the extrinsic curvature playing the role of kinetic energy and the negative of the intrinsic curvature that of potential energy.

The form (2.6) is manifestly invariant under 3-dimensional general coordinate transformations. Precisely for this reason it differs from the Lagrangian of ordinary field theories, for the ${}^{(3)}R$ term of its integrand contains linearly occurring second spatial derivatives of the field variables. With an ordinary field theory in an infinite universe this would be of no significance. The usual assumption that the field vanishes outside some arbitrarily large but finite spatial domain permits linearly occurring second derivatives to be eliminated by partial integration without affecting either the dynamical equations or the canonical definition of energy. In the case of gravity, however, the field never vanishes outside a finite domain unless space-time is flat, and although such a partial integration leaves the dynamical equations unaffected it does change the definition of energy. It is easy to verify, in fact, that it subtracts from the Lagrangian (2.6) a surface integral E_∞ given by

$$E_\infty = \int_\infty \alpha\gamma^{1/2}\gamma^{ij}(\gamma_{ik,j} - \gamma_{ij,k})\,dS^k, \quad (2.7)$$

and hence adds a corresponding quantity to the canonical energy. In an asymptotically flat world it is always possible to find an asymptotically Minkowskian reference frame in which α, β_i, and γ_{ij} take the static Schwarzschild forms

$$\alpha \underset{r\to\infty}{\longrightarrow} 1 - \frac{M}{16\pi r}, \quad \beta_i \underset{r\to\infty}{\longrightarrow} 0, \quad \gamma_{ij} \underset{r\to\infty}{\longrightarrow} \delta_{ij} + \frac{M}{8\pi}\frac{x^i x^j}{r^3}, \quad (2.8)$$

where $r^2 \equiv x^i x^i$ and M is the effective gravitational mass of the field distribution. Substitution of (2.8) into (2.7) yields

$$E_\infty = M. \quad (2.9)$$

It is to be noted that the removal of E_∞ from the Lagrangian does not correspond to a mere redefinition of the energy zero point. E_∞ is not a fixed constant but depends on the state of the field. In fact it is *the* energy, for as we shall see presently the canonical "energy" based on (2.6) always vanishes. (Indeed, E_∞ is the energy even when other fields are present.) Since neither (2.6) nor (2.7) have any explicit dependence on x^0, the quantity E_∞ is conserved. General relativity is unique among field theories in that its energy may always be expressed as a surface integral. This was the source of Bergmann's hope to use gravity as a regulator, but it is also a source of difficulties. We note in particular that the surface integral vanishes for a closed finite world. It is only for infinite asymptotically flat worlds that the energy concept has meaning.

3. THE CONSTRAINTS

The momenta conjugate to α, β_i, and γ_{ij} will be denoted by π, π^i, and π^{ij}, respectively. They have the explicit forms

$$\pi = \frac{\delta L}{\delta\alpha_{,0}} = 0, \quad (3.1)$$

$$\pi^i = \frac{\delta L}{\delta\beta_{i,0}} = 0, \quad (3.2)$$

$$\pi^{ij} = \frac{\delta L}{\delta\gamma_{ij,0}} = -\gamma^{1/2}(K^{ij} - \gamma^{ij}K), \quad (3.3)$$

Eqs. (3.1) and (3.2) being known as the *primary constraints*. The primary constraints are purely formal statements, which express the fact that the Lagrangian (2.6) is independent of the "velocities" $\alpha_{,0}$ and $\beta_{i,0}$.[14]

[14] Failure to bring the Lagrangian into the form (2.6) was responsible for the difficulties originally encountered with the primary constraints.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

These "velocities" are arbitrary and cannot be re-expressed in terms of momenta. They therefore cannot be removed from the Hamiltonian, which, with the aid of (3.3), takes the form

$$H=\int(\pi\alpha_{,0}+\pi^i\beta_{i,0}+\pi^{ij}\gamma_{ij,0})d^3x-L$$

$$=\int(\pi\alpha_{,0}+\pi^i\beta_{i,0}+\alpha\mathcal{H}+\beta_i\mathcal{X}^i)d^3x, \tag{3.4}$$

where

$$\mathcal{H}\equiv\tfrac{1}{2}\gamma^{-1/2}(\gamma_{ik}\gamma_{jl}+\gamma_{il}\gamma_{jk}-\gamma_{ij}\gamma_{kl})\pi^{ij}\pi^{kl}-\gamma^{1/2}\,{}^{(3)}R \tag{3.5a}$$

$$\equiv\gamma^{1/2}(K_{ij}K^{ij}-K^2-{}^{(3)}R), \tag{3.5b}$$

$$\mathcal{X}^i\equiv-2\pi^{ij}{}_{;j}\equiv-2\pi^{ij}{}_{,j}-\gamma^{il}(2\gamma_{jl,k}-\gamma_{jk,l})\pi^{jk}. \tag{3.6}$$

The momenta, as well as $\mathcal{H}$ and $\mathcal{X}^i$, are all 3-densities of unit weight.

It is not hard to show that Einstein's empty-space field equations may be obtained by taking the Poisson bracket of the various dynamical variables with the Hamiltonian (3.4) and then imposing Eqs. (3.1) and (3.2) as external constraints. Since the undermined "velocities" $\alpha_{,0}$ and $\beta_{i,0}$ are multiplied in (3.4) by π and π^i, their Poisson brackets with anything may be ignored. If desired, one can always assign definite values to α and β_i which may be purely numerical or may depend on the γ_{ij} and π^{ij}. Each choice corresponds to the imposition of certain conditions on the space-time coordinates. For example, one may choose

$$\alpha=1,\quad\beta_i=0, \tag{3.7a}$$

which reduces the Hamiltonian to

$$H=\int\mathcal{H}d^3x. \tag{3.7b}$$

Another favorite choice is

$$K\equiv\tfrac{1}{2}\gamma^{-1/2}\gamma_{ij}\pi^{ij}=0,\quad(\gamma^{1/2}\gamma^{ij})_{,j}=0, \tag{3.8a}$$

which corresponds to the requirement that the volume of every hypersurface $x^0=$ constant be stationary under small timelike deformations,[15] and that the spatial coordinates in each hypersurface be harmonic. To obtain the explicit forms of the conditions which Eqs. (3.8a) impose upon α and β_i, one notes that these equations imply the vanishing not only of their left-hand sides but of all their space-time derivatives as well. Taking the Poisson brackets of K and $(\gamma^{1/2}\gamma^{ij})_{,j}$ with the Hamiltonian (3.4) one finds the conditions

$$\alpha_{,i}{}^{,i}-{}^{(3)}R\alpha=0,$$
$$[\gamma^{1/2}(\beta^i{}_{,j}+\beta^j{}_{,i}-\gamma^{ij}\beta^k{}_{,k})+2\alpha\pi^{ij}]_{,j}=0. \tag{3.8b}$$

In an infinite asymptotically flat world these equations, which are of the elliptic type, may be solved subject to the boundary conditions $\alpha\to1$, $\beta_i\to0$ at infinity in asymptotically Minkowskian coordinates. In a finite world of nonvanishing curvature, however, they usually possess either no physically admissible solutions, i.e., solutions for which α remains everywhere positive, or no solutions at all. Since the Laplace-Beltrami operator has a negative spectrum, the first equation, for example, cannot be solved in a 3-sphere.

Conditions of the above type correspond merely to restrictions on the coordinates and have no physical content. There exist conditions of yet another type which actually restrict the dynamical freedom of the field and which hold regardless of whether a specific choice has been made for α and β_i or not. These are obtained by noting that since the primary constraints hold for all time, the x^0 derivatives of π and π^i must vanish. Stating this in the form of a Poisson bracket with H, one arrives immediately at the so-called *secondary or dynamical constraints*:

$$\mathcal{H}=0, \tag{3.9}$$

$$\mathcal{X}^i=0. \tag{3.10}$$

Equation (3.9) will be called the *Hamiltonian constraint*[16] in virtue of the structure of the function $\mathcal{H}$, which appears in (3.5b) as the difference of the extrinsic and intrinsic curvatures, in analogy with the classic form of the Hamiltonian as the sum of the kinetic and potential energies. This Hamiltonian, however, vanishes, as does indeed the total Hamiltonian (3.4). That is to say, in any "Ricci-flat" space-time (i.e., one satisfying Einstein's empty-space equations) the extrinsic and intrinsic curvatures of any hypersurface are equal. As has been emphasized by Wheeler,[11] the converse of this theorem is also true, namely, if $\mathcal{H}$ vanishes over every hypersurface then space-time is Ricci-flat. This suggests, as will be verified later, that it is the Hamiltonian constraint which provides the essential description of the "intrinsic" (i.e., coordinate-independent) dynamics of the gravitational field.

4. QUANTIZATION, CONSISTENCY OF THE CONSTRAINTS, FACTOR ORDERING

In the quantum theory, Poisson brackets become commutators. This means that the constraint Eqs. (3.1), (3.2), (3.9), and (3.10) cannot become operator equations, for otherwise the Hamiltonian (3.4) would yield no dynamics at all, extrinsic or intrinsic. Instead they become conditions on the state vector Ψ[5,9]:

$$\pi\Psi=0, \tag{4.1}$$

$$\pi^i\Psi=0, \tag{4.2}$$

$$\mathcal{H}\Psi=0, \tag{4.3}$$

$$\mathcal{X}^i\Psi=0. \tag{4.4}$$

[15] If 3-space is infinite this applies to the volume inside every finite domain.

[16] All four constraints (3.9), (3.10) are sometimes referred to as "Hamiltonian constraints." We prefer to reserve the terminology for this particularly important constraint.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

These quantum constraints are often a source of puzzlement and confusion. Consider the equation

$$\gamma_{ij}(x^0,\mathbf{x}) = e^{iHx^0}\gamma_{ij}(0,\mathbf{x})e^{-iHx^0}, \quad \mathbf{x}\equiv(x^1,x^2,x^3) \quad (4.5)$$

which is the quantum-mechanical relation expressing the field operator γ_{ij} on an arbitrary hypersurface in terms of the corresponding operator on the hypersurface $x^0=0$. Suppose we choose α and β_i as in Eq. (3.7a). Then Eq. (4.3) and its conjugate imply[17]

$$H\Psi=0, \quad \Psi^\dagger H=0, \quad (4.6)$$

and hence

$$\Psi^\dagger\gamma_{ij}(x^0,\mathbf{x})\Psi=\Psi^\dagger\gamma_{ij}(0,\mathbf{x})\Psi. \quad (4.7)$$

A similar result holds for any other field operator or product of field operators. Since the statistical results of any set of observations are ultimately expressible in terms of expectation values, one therefore comes to the conclusion that nothing ever happens in quantum gravidynamics, that the quantum theory can never yield anything but a static picture of the world.[18]

To see what is wrong with this conclusion one must examine the behavior of H, or more precisely $\mathcal{H}$, at infinity. In an infinite asymptotically flat world the field disperses ultimately to a state of infinite weakness. In the asymptotic region $\mathcal{H}$ therefore tends to its dominant linear term $\gamma_{ii,jj}-\gamma_{ij,ij}=0$,which is the well-known fourth constraint of linearized gravity theory.[10] This term is the asymptotic limit of the term which is removed from the integrand of H by the partial integration discussed in Sec. 2, and which gives rise to the surface integral (2.7). In the linearized theory, however, it becomes a constraint which has no relation to the total energy. Therefore if the full theory is to be applicable not only in the nonlinear region but also at infinity where the linear theory holds sway, it must make use of the Hamiltonian

$$H_\infty \equiv H+E_\infty, \quad (4.8)$$

which results from the partial integration. The integrand of this Hamiltonian reduces, in the asymptotic region, to an expression *quadratic* in the γ's and π's, namely, the usual integrand of the linearized theory.

It follows that in an infinite asymptotically flat world Eq. (4.5) should be replaced by

$$\gamma_{ij}(x^0,\mathbf{x}) = e^{iH_\infty x^0}\gamma_{ij}(0,\mathbf{x})e^{-iH_\infty x^0}. \quad (4.9)$$

Even with this replacement, however, the appearance of the world is still static whenever Ψ is an eigenstate of energy-momentum. To obtain nonstatic behavior one must construct *wave packets*, by superposing many different momenta. But this is precisely what one wants to do in order to provide S-matrix theory, for example, with a rigorous foundation and insure that the field really does disperse ultimately to a state of infinite weakness.

[17] $\mathcal{H}$ is assumed to be ordered in an Hermitian fashion.
[18] Cf. A. Komar, Phys. Rev. 153, 1385 (1967).

Although the above discussion makes use of the coordinate system defined by Eqs. (3.7a), the same problems arise in any other asymptotically Minkowskian coordinate system, and the same conclusions apply. To the extent that we can ignore the possible lack of commutativity of α and β_i with $\mathcal{H}$ and x^i in the construction of an Hermitian Hamiltonian, the same apparent static behavior of the field will occur whenever we incorrectly use H instead of H_∞ in Eq. (4.9).

It should be noted that coordinate conditions such as (3.7a) and (3.8a) are *operator* equations and not constraints on the state vector.[19] (This follows from the complete arbitrariness of α and β_i in the classical theory.) On the other hand, equations such as (3.8a), which hold only when α and β_i are suitably restricted, are *not* operator equations. Indeed, they are not even constraints, but become instead *expectation-value equations*

$$\Psi^\dagger K\Psi=0, \quad \Psi^\dagger(\gamma^{1/2}\gamma^{ij})_{,j}\Psi=0, \quad (3.8c)$$

which hold for all values of x^0 provided they hold at some initial instant and Eqs. (3.8b) are satisfied. They do not hold in all permissible states merely in virtue of (3.8b).

Although we know that the physical content of the classical theory is unaffected by the choice of coordinates, it is not so easy to prove, using the canonical theory, that the results of a calculation of some physical *quantum* amplitude is independent of the choice of coordinates. It is not enough merely to know, for example, that two different coordinate systems both take the Minkowskian form $\alpha\to 1$, $\beta_i\to 0$, $\gamma_{ij}\to\delta_{ij}$ at infinity, in order to conclude that the physical S matrix remains unchanged under the transformation from one system to the other, for the operator ΔH, which represents the change in the Hamiltonian in passing from one system to the other, produces effects which propagate to infinity. In order to prove invariance of the S matrix under coordinate transformations (including q-number coordinate transformations), one would have to show that ΔH affects only the nonphysical field modes at infinity. The obstacle to such a demonstration is the lack of commutativity of the operators appearing in the dynamical equations, particularly when α and β_i depend nonlocally on γ_{ij} and π^{ij}. Although noncommutativity has no effect on the scattering amplitudes in lowest order, it plays havoc with the radiative corrections. For the study of radiative corrections a manifestly covariant theory is almost essential. In the following paper of this series the theory of gravitational radiative

[19] There is an alternative approach to the quantum theory of gravity which makes use of an action functional which is not coordinate-invariant and which generates no primary or secondary constraints. In this approach the constraints must be imposed from the outside. They take the form of coordinate conditions whose form is not arbitrary but is determined by the action functional itself. In this case the coordinate conditions *are* constraints on the state vector. This is the approach which has been followed, for example, by Gupta [S. N. Gupta, in *Recent Developments in General Relativity* (Pergamon Press, Inc., New York, 1962)].

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

corrections will be displayed in all its complexity, and the S matrix in the manifestly covariant theory will be proved to be fully coordinate invariant. This result has not yet been proved in the canonical theory, and for this reason we shall include little further discussion of the case of infinite asymptotically flat worlds in this paper, but will concentrate henceforth on finite worlds.

In the finite case there is no distinction between H and H_∞, and hence we must face up anew to the difficulties posed by Eq. (4.7). The following procedure will be adopted: Instead of regarding this equation as implying that the universe is static we shall interpret it as informing us that the coordinate labels x^μ are really irrelevant. Physical significance can be ascribed only to the intrinsic dynamics of the world, and for the description of this we need some kind of intrinsic coordinatization based either on the geometry or the contents of the universe. In the case of infinite asymptotically flat worlds the Minkowski coordinates at infinity have independent physical relevance as preferred coordinates (up to a Lorentz transformation) based on an *a priori assumed* isometry group (the Poincaré group) for the asymptotic region. One may say that they are intrinsically determined by an implicit laboratory or observer at infinity, and that the constraints serve merely to eliminate the nonphysical modes from the field. In the case of finite worlds, however, the constraints are everything; they and they alone must yield the complete quantum-mechanical description of the world geometry. One of our tasks in the remainder of this paper will be to try to convince the reader that the equations of constraint really do *saturate* the theory, that nothing else is needed.

We must first establish the fact that the constraints are consistent with each other, and this raises some issues of factor ordering.[20] Unfortunately, general agreement has not yet been reached on how to resolve these issues, and hence the proposals which follow must be regarded as tentative. We emphasize, however, our view that the factor-ordering question is not very important to the theory as a whole, and should in no case be permitted to impede attempts to apply the theory to concrete problems. It arises in every local-field theory possessing nontrivial spectral functions, and bears mainly on problems of interpreting divergences. The latter are always resolved by symmetry arguments or by removing infinities from divergent integrals in an invariant way. How such procedures operate in the case of gravity will appear in the papers devoted to the manifestly covariant theory, where questions of factor ordering will again be discussed.

Consistency of the constraints is established if it can be shown that commutators of the constraints lead to no new constraints. The basic commutation relations of the canonical variables themselves are

$$[\alpha, \pi'] = i\delta(\mathbf{x},\mathbf{x}'), \quad [\beta_{i'}, \pi^{j'}] = i\delta_{i'}{}^{j'},$$
$$[\gamma_{ij}, \pi^{k'l'}] = i\delta_{ij}{}^{k'l'};$$
$$\text{all other commutators vanish;} \tag{4.10}$$

in which a notation is employed which emphasizes the bitensor transformation character of the quantities on the right, with primes being used, either on indices or on the variables themselves, to distinguish different points of 3-space. Here $\delta(\mathbf{x},\mathbf{x}')$ denotes the 3-dimensional δ function, and

$$\delta_{i'}{}^{j'} \equiv \delta_i{}^j\delta(\mathbf{x},\mathbf{x}'), \quad \delta_{ij}{}^{k'l'} \equiv \delta_{ij}{}^{kl}\delta(\mathbf{x},\mathbf{x}'),$$
$$\delta_{ij}{}^{kl} \equiv \tfrac{1}{2}(\delta_i{}^k\delta_j{}^l + \delta_i{}^l\delta_j{}^k). \tag{4.11}$$

(The δ function will ordinarily be viewed as a bidensity of zero weight at its first argument and of unit weight at its second.)

The primary constraints evidently give no trouble, since they commute with each other and with the secondary constraints. We therefore turn to the latter and look first at the $\mathcal{X}$ constraints. These will be taken precisely as written in Eq. (3.6), with the momentum factor π^{jk} standing to the right. However, the index will be lowered by defining

$$\mathcal{X}_i \equiv \gamma_{ij}\mathcal{X}^j, \tag{4.12}$$

which, since γ_{ij} stands to the left, yields an alternative form for Eq. (4.4):

$$\mathcal{X}_i\Psi = 0. \tag{4.13}$$

$\mathcal{X}_i$ has the important property of being homogeneous bilinear in the γ_{ij} and the π^{ij}, with the γ's to the left and the π's to the right. Therefore its commutator with any other $\mathcal{X}_{j'}$ has the same property. To compute this commutator it is helpful first to compute the following:

$$\left[\gamma_{ij}, i\int \mathcal{X}_{k'}\delta\xi^{k'}d^3x'\right]$$
$$= -\gamma_{ij,k}\delta\xi^k - \gamma_{kj}\delta\xi^k{}_{,i} - \gamma_{ik}\delta\xi^k{}_{,j}, \tag{4.14}$$

$$\left[\pi^{ij}, i\int \mathcal{X}_{k'}\delta\xi^{k'}d^3x'\right]$$
$$= -(\pi^{ij}\delta\xi^k)_{,k} + \pi^{kj}\delta\xi^i{}_{,k} + \pi^{ik}\delta\xi^j{}_{,k}, \tag{4.15}$$

which reveal the $\mathcal{X}$'s as generators of 3-dimensional coordinate transformations. Under the infinitesimal coordinate transformation $\bar{x}^i = x^i + \delta\xi^i$, the change in any function of the γ_{ij}, π^{ij} and their derivatives is given by commutation with $i\int \mathcal{X}_i\delta\xi^i d^3x$, provided the function has no explicit dependence on x. From this it follows at once that

$$[\mathcal{X}_i, \mathcal{X}_{j'}] = -i\int \mathcal{X}_{k''}C^{k''}{}_{ij'}d^3x'', \tag{4.16}$$

[20] See, for example, J. L. Anderson, in *Proceedings of the 1962 Eastern Theoretical Conference*, edited by M. E. Rose (Gordon and Breach Science Publishers, Inc., New York, 1963), p. 387. See also J. Schwinger, Phys. Rev. 130, 1253 (1963); 132, 1317 (1963).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where the c's are the structure constants of the general coordinate transformation group:

$$c^{k''}{}_{i'j'} \equiv \delta^{k''}{}_{i,l''}\delta^{l''}{}_{j'} - \delta^{k''}{}_{j',l''}\delta^{l''}{}_{i'}. \tag{4.17}$$

The same observations, combined with the fact that $\mathcal{H}$ is a scalar density, yield the formula

$$[\chi_i, \mathcal{H}'] = i\mathcal{H}\delta_{,i}(\mathbf{x},\mathbf{x}'). \tag{4.18}$$

Again the ordering of factors remains the same on both sides of the equation. The only term of $\mathcal{H}$ which might lead to difficulty is the one quadratic in the momenta. But all of the factors which appear in this term have homogeneous linear transformation laws under the 3-dimensional coordinate transformation group and hence remain undisturbed in position when commuted with χ_i. Thus, the commutators (4.16) and (4.18) yield no new constraints, and the choice of factor ordering for $\mathcal{H}$ is so far arbitrary.

Now note that all of the above results could have been obtained equally well had the opposite ordering been chosen for χ_i, with the π's standing to the left and the γ's to the right. The difference between the two choices for χ_i therefore commutes with everything and is evidently a c number. It is a c number, moreover, with definite transformation properties; namely, it is a covariant 3-vector density. From this we may conclude that it can only be zero, for otherwise 3-space would contain a preferred direction quite independently of any geometry which may be imposed on it. The reasonableness of this conclusion also follows from a straightforward formal computation of the difference between the two χ's, which yields derivatives of δ functions with coincident arguments. Any ordering may therefore be chosen for χ_i, and if γ_{ij} and π^{ij} are Hermitian so is χ_i.

The same conclusions do not automatically hold for χ^i, since the difference between two orderings for it involves an undifferentiated δ function. Let us therefore see what we can say about the formal symbol $\delta(\mathbf{x},\mathbf{x})$. Consider the third commutator in (4.10). If we set $\mathbf{x}'=\mathbf{x}$ and contract all the indices, we obtain

$$(\gamma_{ij}\pi^{ij} - \pi^{ij}\gamma_{ij}) = 6i\delta(\mathbf{x},\mathbf{x}). \tag{4.19}$$

The quantity on the right is certainly a c number. Therefore we may write

$$[6i\delta(\mathbf{x},\mathbf{x}), i\int \chi_{k'}\delta\xi^{k'}d^3x'] = 0. \tag{4.20}$$

On the other hand, if we apply the same commutator to the left we obtain

$$[(\gamma_{ij}\pi^{ij} - \pi^{ij}\gamma_{ij}), i\int \chi_{k'}\delta\xi^{k'}d^3x']$$

$$= -[(\gamma_{ij}\pi^{ij} - \pi^{ij}\gamma_{ij})\delta\xi^k]_{,k} = -6i[\delta(\mathbf{x},\mathbf{x})\delta\xi^k]_{,k}. \tag{4.21}$$

Equating the two results we find

$$[\delta(\mathbf{x},\mathbf{x})\delta\xi^i]_{,i} = 0. \tag{4.22}$$

This equation must hold for arbitrary $\delta\xi^i$. Therefore, although most people would say that $\delta(\mathbf{x},\mathbf{x})$ is infinite, we see that it is actually zero.

In order to understand how this formal result can be consistent with the rest of the theory one must first note that Eqs. (4.3) and (4.13) are really abbreviations for the correct forms

$$\int \mathcal{H}\xi d^3x\ \Psi = 0 \quad \text{for all } \xi, \tag{4.23}$$

$$\int \chi_i\xi^i d^3x\ \Psi = 0 \quad \text{for all } \xi^i, \tag{4.24}$$

where ξ and ξ^i are arbitrary but *smooth* c-number weight functions. The problem of taking commutators of field quantities at the same space-time point therefore never arises with pairs of constraints but only in connection with the definition of the functions $\mathcal{H}$ and χ_i themselves. This means that the δ function may, without inconsistency, be thought of as the limit of a sequence of successively narrower *twin-peaked* functions, all of which are smooth, have unit integral, and vanish at the point $\mathbf{x}'=\mathbf{x}$ in the valley between the peaks. An example of such a function in one dimension would be $\delta(x) = \lim(2\pi)^{-1}[f_\epsilon(x-\sqrt{\epsilon}) + f_\epsilon(x+\sqrt{\epsilon}) - 2f_\epsilon(x)/(1+\epsilon)]$, where $f_\epsilon(x) \equiv \epsilon(x^2+\epsilon^2)^{-1}$. In an infinite world, passage to the limit $\epsilon \to 0$ would correspond to the usual cutoff going to infinity in momentum space, while maintenance of the valley at $\mathbf{x}'=\mathbf{x}$ would yield a particular regularization of the resulting divergences. The answer to the question whether or not this regularization is equivalent to the quite different procedures which will prove useful in the manifestly covariant theory must await a demonstration of how to derive one theory from the other. In the meantime we shall in *this* paper simply adopt it as a rule that any two field operators taken at the same space-time point commute. The consistency question for the constratints then reduces to that of the classical theory.

There remains to be considered only the commutator $[\mathcal{H},\mathcal{H}']$. At first sight it might be thought that the commutator of the two quadratic-in-the-momenta terms, one from $\mathcal{H}$ and the other from $\mathcal{H}'$, leads to difficulties. However, these terms contain no derivatives (of the γ's or π's) with respect to the 3-space coordinates and hence they commute. Since the terms $\gamma^{1/2}\,{}^{(3)}R$ and $\gamma'^{1/2}\,{}^{(3)}R'$ contain no momenta, they likewise commute. The only commutators which remain are the cross commutators, and these can be evaluated by judicious use of the variational formula

$$\delta(\gamma^{1/2}\,{}^{(3)}R) = \gamma^{1/2}\gamma^{ik}\gamma^{jl}(\delta\gamma_{ik,jl} - \delta\gamma_{ij,kl}) - \gamma^{1/2}({}^{(3)}R^{ij} - \tfrac{1}{2}\gamma^{ij}\,{}^{(3)}R)\delta\gamma_{ij}. \tag{4.25}$$

The final result is

$$[\mathcal{H},\mathcal{H}'] = 2i\mathcal{H}^i\delta_{,i}(\mathbf{x},\mathbf{x}') + i\mathcal{H}^i{}_{,i}\delta(\mathbf{x},\mathbf{x}'), \tag{4.26a}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

or, more correctly,

$$\left[\int \mathcal{H}\xi_1 d^3x, \int \mathcal{H}\xi_2 d^3x\right] = i\int \chi^i(\xi_1\xi_{2,i} - \xi_{1,i}\xi_2)d^3x. \quad (4.26b)$$

If we were still concerned about the order of factors we would find that a symmetric (Hermitian) ordering for $\mathcal{H}$ would yield a symmetric ordering for χ^i in (4.26), namely $\chi^i = \frac{1}{2}\{\gamma^{ij}, \chi_j\}$, and the problem at issue would then be to evaluate the commutator $[\gamma^{ij}, \chi_j]$. From our present point of view this commutator vanishes, and consistency is maintained.[21]

5. THE FUNCTIONAL WAVE EQUATION AND THE STATE-FUNCTIONAL DOMAIN MANIFOLD

Further analysis of the canonical theory requires the introduction of a specific representation for the quantum states. Wheeler[11] has chosen for this purpose what may be called the *metric representation*, in which Ψ becomes a functional of the metric components $g_{\mu\nu}$, and the momenta become functional differential operators:

$$\pi = \frac{\delta}{i\delta\alpha}, \quad \pi^i = \frac{\delta}{i\delta\beta_i}, \quad \pi^{ij} = \frac{\delta}{i\delta\gamma_{ij}}. \quad (5.1)$$

The primary constraints tell us that Wheeler's Ψ depends only on the γ's. We shall indicate this, for the present, by writing Ψ in the form $\Psi[\gamma]$. (Since we are working in a closed finite world, it would be meaningless to include also a dependence on x^0.)

Consider now the χ constraints. In the metric representation these take the form

$$2i(\delta\Psi[\gamma]/\delta\gamma_{ij})_{,j} = 0, \quad (5.2)$$

which are the necessary and sufficient conditions that $\Psi[\gamma]$ be an invariant under coordinate transformations. In a finite world this means that Ψ depends only on the geometry of 3-space. One possible way to express this dependence would be to regard Ψ as a function of a discrete infinity of variables, namely all the independent invariants, beginning with $\int\gamma^{1/2}d^3x$, $\int\gamma^{1/2}\,{}^{(3)}R d^3x$, $\int\gamma^{1/2}\,{}^{(3)}R^2 d^3x$, etc., which can be constructed out of products of the Riemann tensor and its covariant derivatives, with the topology of 3-space itself being separately specified.

Higgs[8] has pointed out that in an infinite world such a characterization of Ψ would be inadequate, for in this case the asymptotic coordinates also play a role. Ψ could instead be represented as a functional of any three of the six coordinate-invariant functions[22]:

$$\varphi^{AB}(\eta) \equiv \int \gamma^{1/2}\gamma^{ij}\zeta^A{}_{,i}\zeta^B{}_{,j}\delta^3(\eta - \zeta(x))d^3x, \quad A, B = 1, 2, 3, \quad (5.3)$$

where the ζ's are scalars satisfying the elliptic differential equation

$$\zeta^A{}_{,i}{}^{,i} = 0, \quad (5.4)$$

with the boundary conditions $\zeta^A \to x^A$ at infinity. The ζ's define a harmonic coordinate system, and Eqs. (5.4) yield $\partial\varphi^{AB}/\partial\eta^B = 0$ as a corollary. If φ^{11}, φ^{12}, φ^{22} are arbitrarily chosen then φ^{13}, φ^{23}, φ^{33} are determined by integrating successively the equations $\partial\varphi^{13}/\partial\eta^3 = -\partial\varphi^{11}/\partial\eta^1 - \partial\varphi^{12}/\partial\eta^2$, $\partial\varphi^{23}/\partial\eta^3 = -\partial\varphi^{12}/\partial\eta^1 - \partial\varphi^{22}/\partial\eta^2$, $\partial\varphi^{33}/\partial\eta^3 = -\partial\varphi^{13}/\partial\eta^1 - \partial\varphi^{23}/\partial\eta^2$. If space-time is asymptotically flat and the coordinates x^i are Minkowskian at infinity, then these equations can be consistently integrated with the asymptotic boundary conditions $\varphi^{AB} \to \delta_{AB}$.

The above example is cited in order to re-emphasize the fundamental difference between finite and infinite worlds. In the finite case we may replace the symbol $\Psi[\gamma]$ by $\Psi[{}^{(3)}\mathcal{G}]$ to display the fact that Ψ depends only on the 3-geometry, denoted here by ${}^{(3)}\mathcal{G}$, and on nothing else, whereas in the infinite case we must write something like $\Psi[{}^{(3)}\mathcal{G}, \mathfrak{L}]$, with $\mathfrak{L}$ symbolizing the surrounding laboratory which determines the asymptotic coordinates (including, in the Schrödinger picture, the coordinate x^0).

We shall denote by $\mathfrak{M}$ the set of all possible 3-geometries which a finite world may possess. The following question will arise: Can a topology be imposed upon $\mathfrak{M}$ which is both meaningful and at the same time useful in the context of the quantum theory of finite worlds? One possibility which suggests itself is to view $\mathfrak{M}$ as an infinite-dimensional vector space whose "points" are discrete sets of invariants mentioned earlier. The topology could be that defined by the Cartesian metric on this space, and the symbol ${}^{(3)}\mathcal{G}$ could be replaced by a set of vector components. In fact, this possibility is not very useful, and although we shall actively pursue the question of assigning a metric, and indeed a pseudo-Riemannian structure, to $\mathfrak{M}$, no advantage will be gained by attempting to make our symbolism more explicit. It will be sufficient simply to keep in mind the idea that $\mathfrak{M}$ is not just a mere set but is actually a

[11] J. Schwinger (Ref. 20) proposes an alternative resolution of the factor ordering problem which, in the notation of the present paper, runs essentially as follows: Replace $\mathcal{H}$ in the Hamiltonian constraint by $(\gamma^{1/2}\mathcal{H})$, where ( ) indicates that the factors are to be placed in some (arbitrary) symmetrical order. Then compute

$$[(\gamma^{1/2}\mathcal{H}), (\gamma^{1/2}\mathcal{H}')] = \tfrac{1}{2}i[(\gamma^{ij}, (\chi_j)) + (\gamma^{ij}, (\chi_{j'}))]\delta_{,i}(x,x') = \tfrac{1}{2}i[(\gamma^{ij}, (\chi_{j'})) + (\gamma^{ij}, (\chi_j))]\delta_{,i}(x,x').$$

Since the commutator

$$[\gamma^{ij}, (\chi_{j'})] = i(\gamma^{ij}\gamma^{kl} + \gamma^{il}\gamma^{kj})\delta_{,j}(x,x')$$

is antisymmetric in x and x', it follows that

$$[\gamma^{ij}, (\chi_{j'})] + [\gamma^{il}, (\chi_j)] = 0,$$

whence

$$[(\gamma^{1/2}\mathcal{H}), (\gamma^{1/2}\mathcal{H}')] = i[\gamma^{ij}(\chi_{j'}) + \gamma^{il}(\chi_j)]\delta_{,i}(x,x')$$

in which the χ's stand to the right. Demonstration of consistency of the other commutators is elementary.

[22] In the Schrödinger picture Ψ would also depend on x^0.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

manifold. Thus we shall say: $\mathfrak{M}$ is the *domain manifold* for the state functional Ψ, and the $^{(3)}\mathcal{G}$ are its "points."

So far nothing has been said about dynamics. The only way in which dynamics can enter the picture is through the Hamiltonian constraint. This now takes the form

$$\left(G_{ijkl}\frac{\delta}{\delta\gamma_{ij}}\frac{\delta}{\delta\gamma_{kl}}+\gamma^{1/2}\,{}^{(3)}R\right)\Psi[{}^{(3)}\mathcal{G}]=0, \qquad (5.5)$$

where

$$G_{ijkl}\equiv\tfrac{1}{2}\gamma^{-1/2}(\gamma_{ik}\gamma_{jl}+\gamma_{il}\gamma_{jk}-\gamma_{ij}\gamma_{kl}). \qquad (5.6)$$

According to our rule of freely commuting field operators taken at the same space-time point, the functional differential operator $\delta/\delta\gamma_{ij}$ must always be understood to give zero when acting on a γ_{kl} at the same point.[23] If it were not for this rule, we might try to regard the first term in the parentheses of (5.5) as a kind of Laplace-Beltrami operator in a 6-dimensional Riemannian manifold having G_{ijkl} as its *contra*variant metric. Although such an interpretation is inappropriate for the operator itself, it *is* nevertheless useful to regard G_{ijkl} as a metric tensor and to study the properties of the manifold which it defines. These properties, which are derived in Appendix A, turn out to be quite interesting.

The manifold in question will be denoted by M. When γ_{ij} is positive definite (as it is for a spacelike hypersurface) M has the hyperbolic signature $-+++++$. A "pure dilation" of γ_{ij} (i.e., multiplication by a multiple of the unit matrix) constitutes a typical "timelike" displacement. It is convenient to introduce the timelike coordinate

$$\zeta\equiv(32/3)^{1/2}\gamma^{1/4} \qquad (5.7)$$

and any five other coordinates ζ^A orthogonal to it. The covariant metric then takes the form

$$\begin{pmatrix}-1 & 0\\ 0 & (3/32)\zeta^2\bar{G}_{AB}\end{pmatrix}, \qquad (5.8)$$

where

$$\bar{G}_{AB}\equiv\mathrm{tr}(\gamma^{-1}\gamma_{,A}\gamma^{-1}\gamma_{,B}), \qquad (5.9)$$

$$\gamma\equiv(\gamma_{ij}). \qquad (5.10)$$

Expression (5.8) reveals M as a set of "nested" 5-dimensional submanifolds, all having the same intrinsic shape and differing only in the scale factor $(3/32)\zeta^2$. The shape is described by the positive-definite metric $\bar{G}_{AB}$ which, since expression (5.9) remains invariant under a dilation of the γ's, is independent of ζ.

The manifold having $\bar{G}_{AB}$ as a metric will be denoted by $\bar{M}$. It is shown in Appendix A that the geodesic equation in $\bar{M}$ takes the form

$$\frac{d^2\gamma}{ds^2}-\frac{d\gamma}{ds}\gamma^{-1}\frac{d\gamma}{ds}=0, \quad \mathrm{tr}\left(\gamma^{-1}\frac{d\gamma}{ds}\right)=0. \qquad (5.11)$$

This has the general solution[24]

$$\gamma(s)=\mathbf{M}^{\sim}e^{\mathbf{N}s}\mathbf{M}, \qquad (5.12)$$

where $\mathbf{M}$ is an arbitrary nonsingular 3×3 matrix and $\mathbf{N}$ is subject only to the restrictions

$$\mathbf{N}^{\sim}=\mathbf{N}, \quad \mathrm{tr}\mathbf{N}=0, \quad \mathrm{tr}\mathbf{N}^2=1, \qquad (5.13)$$

the last of which guarantees that s is the arc length. Since $e^{\mathbf{N}s}$ is analytic for all values of s, $\bar{M}$ is geodesically complete. It is not difficult to show that any two points of $\bar{M}$ may be joined by a unique geodesic and that if the two points are represented by symmetric matrices γ_1 and γ_2 having the same determinant then their distance of separation is $\{\mathrm{tr}[\ln(\gamma_1^{-1}\gamma_2)]^2\}^{1/2}$. The manifold $\bar{M}$ is evidently noncompact and diffeomorphic to Euclidean 5-space.

By straightforward computation one may verify that the Riemann and Ricci tensors of $\bar{M}$ have the respective forms

$$\bar{R}_{ABCD}=\mathrm{tr}[\gamma^{-1}\gamma_{,D}\gamma^{-1}\gamma_{,C}\gamma^{-1}\times(\gamma_{,A}\gamma^{-1}\gamma_{,B}-\gamma_{,B}\gamma^{-1}\gamma_{,A})], \qquad (5.14)$$

$$\bar{R}_{AB}=-\tfrac{3}{2}\bar{G}_{AB}. \qquad (5.15)$$

From the latter it follows that $\bar{M}$ is an "Einstein space" of constant negative Gaussian curvature. It is furthermore not difficult to show that the Riemann tensor (5.14) has vanishing covariant derivative, which implies that $\bar{M}$ is, in fact, *a symmetric space*[25] with a certain group structure. The group structure may be deduced from the observation that the transformation

$$\gamma'=\mathbf{L}^{\sim}\gamma\mathbf{L}, \qquad (5.16)$$

where $\mathbf{L}$ is an arbitrary constant nonsingular 3×3 matrix, leaves the metric (5.9) unchanged. The full linear group in three dimensions therefore acts isometrically on $\bar{M}$. Because of the dilation invariance of the points of $\bar{M}$, however, it is only the simple Lie Group $SL(3,R)$ which acts *effectively* on it. It is easily verified that $SL(3,R)$ acts transitively on $\bar{M}$ and, moreover, that the *isotropy subgroup*[25] at any point is isomorphic to $SO(3)$. $\bar{M}$ may therefore be identified as the coset space

$$\bar{M}=SL(3,R)/SO(3). \qquad (5.17)$$

Although the manifold $\bar{M}$ is geodesically complete, the manifold M is not. It is shown in Appendix A that all geodesics in M ultimately hit a *frontier* of infinite

[23] There is nothing automatically pathological, however, about having two functional derivatives acting at the same point, as in (5.5). For example, if $I\equiv\frac{1}{2}\int dx\int dx'\varphi(x)K(x,x')\varphi(x')$, where φ is an arbitrary function and K is a fixed kernel, then $\delta^2I/\delta\varphi(x)\delta\varphi(x)=K(x,x)$. Pathology occurs only if $K(x,x')$ is singular at $x'=x$.

[24] The tilde "$\sim$" denotes the transpose. All matrices are assumed real.

[25] See, for example, S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., New York, 1962). The author is indebted to Professor Helgason for enlightment as to the group structure of $\bar{M}$.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

curvature. "Timelike" and null geodesics hit it at one end; "spacelike" geodesics hit it at both ends. This frontier, which will be denoted by F, is located at $\zeta=0$, as may be inferred from the readily computed curvature scalar

$$^{(6)}R=-60/\zeta^2. \tag{5.18}$$

A question now arises as to what estent the Riemannian structure of M may be regarded as imposing a structure on $\mathfrak{M}$ by way of the Hamiltonian constraint. Without attempting to answer this question directly, we may point out certain very suggestive features of the theory. First of all, the existence of the timelike coordinate ζ in M suggests that a corresponding "intrinsic time" exists in $\mathfrak{M}$ and that the Hamiltonian constraint does indeed have dynamical content. This idea is given support by the following considerations: The specification of a given 3-geometry requires the assignment of essentially 3 independent quantities at each point of 3-space. If we regard the usual enumeration of the degrees of freedom possessed by the gravitational field, namely *two* for every point of 3-space, as being valid in a finite world, this leaves one quantity per 3-space point to play the role of intrinsic time. Baierlein, Sharp, and Wheeler[11,25] have shown in the classical theory that if the intrinsic geometry is given on any two hypersurfaces then, except in certain singular cases, the geometry of the entire space-time manifold, *and hence the absolute time lapse between the two hypersurfaces*, is determined. Moreover, it is determined solely by the constraints. Analogously, the quantum theory is completely determined by the transformation functional $\langle {}^{(3)}\mathcal{G}'|{}^{(3)}\mathcal{G}''\rangle$, where $|{}^{(3)}\mathcal{G}\rangle$ denotes that state of the gravitational field for which there exists at least one hypersurface having an infinitely precise geometry ${}^{(3)}\mathcal{G}$. Wheeler[11] has emphasized the imporatnce of the two-hypersurface formulation of gravidynamics (or "geometrodynamics" as he calls it) and has suggested the use of the Feynman sum-over-histories method to compute the transformation functional.[27]

Another suggestive feature of the theory is the following. Because of the hyperbolic character of M the Hamiltonian constraint (5.5) resembles a Klein-Gordon equation, with $-\gamma^{1/2}\,{}^{(3)}R$ playing the role of the mass-squared term. An important difference, however, is that ${}^{(3)}R$ can be either positive or negative, and hence the "wave" propagation of the state functional is not confined to timelike directions.

In spite of this difference the analogy with the Klein-Gordon theory suggests the following definition for the quantum-mechanical inner product of two states Ψ_a and Ψ_b:

$$(\Psi_b,\Psi_a)=Z\int_\Sigma \Psi_b^*[{}^{(3)}\mathcal{G}]$$

$$\times\prod_x\left(d\Sigma^{ij}G_{ijkl}\frac{\vec{\delta}}{i\delta\gamma_{kl}}-\frac{\dot{\delta}}{i\delta\gamma_{kl}}G_{ijkl}d\Sigma^{ij}\right)\Psi_a[{}^{(3)}\mathcal{G}]. \tag{5.19}$$

The infinite product, which arises because (5.5) is really not just one equation but ∞^3 equations, is here taken over all the points of 3-space, and is to be understood in a formal sense as representing the result of a limiting process based on a sequence of lattices in 3-space, each lattice requiring the introduction of a corresponding normalizing constant Z. The symbol Σ denotes the topological product of a set of 5-dimensional M-hypersurfaces $\Sigma(x)$ (one chosen at each point of 3-space), the $d\Sigma^{ij}$ being their directed surface elements. It is an immediate consequence of the Hamiltonian constraint that this inner product is independent of the choice of $\Sigma(x)$'s provided some kind of appropriate boundary conditions are satisfied at the "edges" of Σ. It is also worth noting that since the G_{ijkl} do not involve any spatially differentiated γ's, the operators standing in the infinite product all commute, and hence no factor-ordering difficulties arise here.

In view of the coordinate invariance of the state functionals the inner product integral (5.19) contains a $3\times\infty^3$-fold redundancy arising from the geometrical indistinguishability of 3-metrics which differ only by coordinate transformations.[28] This produces a divergence which must be formally absorbed into the normalization constant Z, and reminds us that $\mathfrak{M}$ is not just the topological product of M with itself over all the points of 3-space, but is a subspace of the latter manifold.

Another difficulty with the definition (5.19) concerns the problem of "negative probability." This problem arises here, just as it does for the Klein-Gordon equation, from the fact that the Hamiltonian constraint involves a second derivative with respect to the "time" coordinate. If the $\Sigma(x)$'s are chosen "spacelike," then the only way to assure positive definiteness of (5.19), when $\Psi_b=\Psi_a$, is to restrict the content of Ψ_a to "positive frequency" components with respect to every "time" coordinate $\zeta(x)$. Restriction to such components, however, implies that Ψ_a vanishes nowhere in the range $-\infty<\zeta<\infty$, and this conflicts with the one-sided character of ζ, namely $\zeta>0$, which follows from the geometrical analysis revealing the existence of a frontier in M at $\zeta=0$. One might hope that an analytic continuation could be performed around $\zeta=0$, but

[25] R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. 126, 1864 (1962). There is nothing mysterious about the existence of a manifold of "time" variables. The same manifold exists in conventional field theory in those formulations which make the state functional depend on an arbitrary spacelike hypersurface.

[27] The sum-over-histories or "functional integral" method has not yet been applied to any "practical" problem of quantum gravidynamics. It will be encountered in heuristic and formal applications in the following papers of this series. Its consistency with the Dirac theory has been demonstrated by Leutwyler [See H. Leutwyler, Phys. Rev. 134, B1155 (1964).]

[28] A coordinate transformation generally produces a change in Σ, but this does not affect the integral.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

whether this would have any physical meaning is unclear. The singularity in the Hamiltonian constraint at $\zeta=0$ is a strong one, as may be seen by rewriting (5.5) in the form

$$\left[-\frac{\delta^2}{\delta\zeta^2}+\frac{(32/3)}{\zeta^2}\bar{G}^{AB}\frac{\delta^2}{\delta\zeta^A\delta\zeta^B}+(3/32)\zeta^2\,{}^{(3)}R\right]\times\Psi[{}^{(3)}\mathcal{G}]=0, \quad (5.20)$$

which makes use of (5.7) and (5.8). The question at issue is whether the frontier in M generates a corresponding barrier in $\mathfrak{M}$ beyond which there is no possibility of extending the state functional. Unfortunately, in the present state of our knowledge no clear-cut answer can be given to this question. Some of the problems which have a bearing on it, however, can be identified. These will now be discussed.

6. THE METRIC OF $\mathfrak{M}$. THE HAMILTON-JACOBI EQUATION AND GRAVITATIONAL COLLAPSE

The most obvious way to approach $\mathfrak{M}$ is through the manifold M^{∞^3}, which is defined formally as the topological product of M with itself over the points of 3-space:

$$M^{\infty^3}\equiv\prod_{\mathbf{x}}M(\mathbf{x}). \quad (6.1)$$

The "points" of M^{∞^3} are the matrix functions $\gamma_{ij}(\mathbf{x})$. For brevity they will be denoted simply by γ. In practice the definition (6.1) must be supplemented by some sort of continuity requirements. For example, γ may be required to be continuous and piecewise differentiable. However, we do not wish to be precise about this here, since as yet no rigorous theory of the role of the manifold $\mathfrak{M}$ in the quantum theory exists. We wish merely to point out some of the issues involved, and to leave the formalism itself as unencumbered as possible. Thus we shall be willing to admit any sort of pathology for γ which we can get away with, i.e., for which some sort of physical interpretation exists, however idealized, which permits γ to be handled in a consistent fashion. For example, geometrical singularities at which the Riemann tensor behaves like a differentiated δ function, or for which integrals like $\int\gamma^{1/2}\,{}^{(3)}R dx$ still exist, will not be excluded *a priori*. In the same spirit, we shall not place any restrictions on coordinate transformations beyond perhaps requiring them to be differentiable (so that tensor transformation laws exist almost everywhere) and one-to-one (so that they form a group). Thus we shall not automatically exclude transformations for which the Jacobian either vanishes or diverges at certain points. The ultimate question will always be: What is the *barrier* beyond which we cannot go? In every case this will probably depend, to some extent at least, on the context, and we do not wish to prejudice the answer in advance.

There is, however, one trivial pathology which may may be avoided without loss of generality, namely, coordinate singularities which arise from the impossibility of covering compact manifolds with a single well-behaved coordinate system. We shall always assume that 3-space is covered with a finite set of overlapping coordinate patches, each of which can be put into one-to-one correspondence with a certain portion of the Cartesian mesh in Euclidean 3-space, and on the boundaries of which the coordinates are held fixed. In addition a set of supplementary connection formulas between patches must be assuemd to hold in the overlap regions. All of this paraphernalia is to be understood as included in the definition (6.1), which means that each function $\gamma_{ij}(\mathbf{x})$ is really a set of functions, one in each coordinate patch, and that the $\mathbf{x}$'s in Eq. (6.1) are to be understood as ranging over all values in all patches.

Now let γ to be a fixed point of M^{∞^3}. Consider the set of all points which may reached from γ by coordinate transformations. This set is known as the *orbit* of γ under the coordinate transformation group and will be denoted by "orb γ." There is a one-to-one correspondence between the orbits in M^{∞^3} and the points of $\mathfrak{M}$. In fact no generality is lost if they are identified:

$$\text{orb}\,\gamma\equiv{}^{(3)}\mathcal{G}. \quad (6.2)$$

Suppose M^{∞^3} is endowed with a metric. (That this is feasible will appear in a moment.) If this metric satisfies a certain condition then it will impose, in a natural way, a metric on $\mathfrak{M}$. The condition is that the coordinate transformation group in 3-space be an *isometry group* of M^{∞^3}, The associated metric in $\mathfrak{M}$ is then obtained by defining the distance between two neighboring orbits to be the shortest distance in M^{∞^3}.

It is shown in Appendix B that the above condition is satisfied if and only if the metric in M^{∞^3} transforms, under 3-dimensional coordinate transformations, contragrediently to the Kronecker product $\gamma_{ij}\gamma_{k'l'}$. This means that the metric in M^{∞^3}, which we shall denote by $\mathcal{G}^{ijk'l'}$, must be a contravariant bitensor density of weight at both x and x'.

There are infinitely many contravariant bitensor densities which can be constructed out of the γ's and which might serve as acceptable metrices for M^{∞^3}. Of these, however, there is only a single one-parameter family which is *local*, i.e., which involves only undifferentiated γ's and for which both $\mathcal{G}^{ijk'l'}$ and its inverse vanish when $x\neq x'$. This family is given by

$$\mathcal{G}^{ijk'l'}=\tfrac{1}{2}\gamma^{1/2}(\gamma^{ik}\gamma^{jl}+\gamma^{il}\gamma^{jk}+\lambda\gamma^{ij}\gamma^{kl})\delta(\mathbf{x},\mathbf{x}'), \quad (6.3)$$

where λ can assume any real value except $-\frac{2}{3}$. If we wished to impose a positive-definite metric on $\mathfrak{M}$, so that we could use, as the condition for the identity of two 3-geometries, the vanishing of the "distance" between them, then the metric of M^{∞^3} itself would have to be positive definite. In the present case this requires $\lambda>-\frac{2}{3}$, the simplest choice being $\lambda=0$. On the other

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

hand, the choice $\lambda=-2$ is the natural choice if we assume that Eq. (6.1) defines not merely a topological product but also a geometrical structure generated by the original metric on M. For then we have

$$\int \mathcal{G}_{ijcd}\mathcal{G}^{cdk'l'}d^3x''=\delta_{ij}{}^{k'l'}, \tag{6.4}$$

where

$$\mathcal{G}_{ijk'l'}\equiv G_{ijkl}\delta(x,x'), \tag{6.5}$$

with G_{ijkl} given by Eq. (5.6). In this case the "arc length" $d\$$ associated with a displacement $d\gamma_{ij}$ in $M^{\infty 3}$ is given by

$$d\$^2=\int d^3x\int d^3x'\,\mathcal{G}^{ijk'l'}d\gamma_{ij}d\gamma_{k'l'}$$

$$=\int \gamma^{1/2}(\gamma^{ik}\gamma^{jl}-\gamma^{ij}\gamma^{kl})d\gamma_{ij}d\gamma_{kl}d^3x. \tag{6.6}$$

Geodesics in $M^{\infty 3}$ are not, in general, geodesics in $\mathfrak{M}$. However, in Appendix B it is shown that if a geodesic in $M^{\infty 3}$ intersects one of the orbits in its path orthogonally then it *is* a geodesic in $\mathfrak{M}$, and, moreover, it intersects every other orbit in its path orthogonally. This means that it is in principle possible to use formula (A69) of the Appendix to determine the distance between two 3-geometries. In practice, of course, the amount of labor involved is formidable, assuming that the 3-geometries are given in the form of two matrix functions $\gamma_1(x)$ and $\gamma_2(x)$. One must integrate expression (A69) over 3-space and then find the minimum of the integral as one of the functions, say $\gamma_1(x)$, is held fixed while the other ranges over the various equivalent forms it can take under coordinate transformations. This means solving the complicated set of nonlinear partial differential equations which result from the corresponding variational principle and which, in effect, yield the coordinate transformation which "lines up" $\gamma_1(x)$ and $\gamma_2(x)$, so that a geodesic from one to the other intersects orbits orthogonally.

Such complications can be avoided if one merely wants to know the distance from a given 3-geometry to the *frontier*.[29] In this case, since the frontier is an extended object and is at different distances—spacelike, timelike, and null—in different directions, it is necessary to specify a direction $d\gamma_{ij}/d\$$ in addition to the 3-geometry itself. Here $\$$ is either the arc length (6.6) or, in the exceptional null case, an affine parameter, and $d\gamma_{ij}/d\$$ must satisfy the starting condition [cf. Eq. (B24)]

$$\gamma^{jk}\left[\left(\frac{d\gamma_{ij}}{d\$}\right)_{,k}-\left(\frac{d\gamma_{jk}}{d\$}\right)_{,i}\right]=0, \tag{6.7}$$

which guarantees that the starting direction will be orthogonal to the starting orbit. The square of the distance in the frontier in the assigned direction is then given by[30]

$$\$^2=\min 2\sigma(\gamma,d\gamma/d\$)\left[G^{ijkl}\frac{d\gamma_{ij}}{d\$}\frac{d\gamma_{kl}}{d\$}\right]^{-1}, \tag{6.8}$$

where G^{ijkl} and σ are defined in Appendix A, Eqs. (A1) and (A63), respectively, and "min" denotes the minimum value over 3-space. The metric tensor at the point (or points) in 3-space at which the minimum occurs will become singular when the "point" γ in $M^{\infty 3}$ has progressed a distance $\$$ along the geodesic. The geodesic can then go no further without changing the signature of a portion of 3-space. The frontier has been reached.

It does not automatically follow that 3-space acquires a *geometrical* singularity at the frontier. However, there are several facts worth noting.

(1) The occurrence of a singular metric cannot be avoided by changing the coordinates as one proceeds along the geodesic. Although it is true that a coordinate transformation can carry one from one point to another in $M^{\infty 3}$ and even, *seemingly*, away from the frontier, yet since expression (6.8) is a scalar, $\$$ remains unchanged. What happens is that the coordinate transformation also changes the direction $d\gamma_{ij}/d\$$. Moreover, the covariance of Eq. (6.7) ensures that the orthogonality of the geodesic to the orbits is left unaffected.

(2) As long as no coordinate transformations are performed while γ is moving along its orthogonal geodesic, the coordinate system in 3-space, if initially nonsingular, will remain nonsingular until the frontier is reached. No such statement can be made for nonorthogonal geodesics, which in some cases follow a circuitous route in $\mathfrak{M}$ from a given ${}^{(3)}\mathcal{G}$ back again to the same ${}^{(3)}\mathcal{G}$, but in a different coordinate system.[31]

(3) It is not necessary that the metric become singular simultaneously at all points of 3-space in order that

[29] By *frontier* we do not necessarily mean *barrier*. It must be repeatedly emphasized that very little is known about the general conditions under which extensions beyond the frontier can be carried out. Here we are defining the frontier to be simply the locus of points γ in $M^{\infty 3}$ for which the matrix $\gamma_{ij}(x)$ has one or more singularity points ($\gamma=0$) in 3-space, regardless of whether or not these singularity points represent real geometrical singularities. A formal definition would be

$$\mathbf{F}=\bigcup_x F(x)\prod_{x'\neq x}M(x'),$$

where $\prod$ and $\bigcup$ denote the topological product and union, respectively, and $F(x)$ is the frontier of $M(x)$.

[30] When $d\gamma_{ij}/d\$$ is a null vector in M, expression (6.8) becomes an indeterminate form 0/0. At such points in 3-space (6.8) may, in view of Eq. (A54) which implies $\$=\text{constant}\times\gamma^{1/2}$, be replaced simply by $\$^2=4(\gamma^{ij}d\gamma_{ij}/d\$)^{-1}$.

[31] In attempting to visualize $M^{\infty 3}$ and $\mathfrak{M}$ it is helpful to have a simpler model in mind. The following is suggested: Let the big manifold be Euclidean 3-space and let the group be rotations about an axis. The orbits are then circles concentric with the axis and at right angles to it, and the orbit manifold is the Euclidean half-plane. A straight line (i.e., geodesic) in the big manifold will be a geodesic in the orbit manifold if it intersects or is parallel to the axis, so that it intersects every circle in its path at right angles. A straight line which is skew to the axis, however, is a hyperbola in the orbit manifold, and a skew line at right angles to the axis returns again to each orbit which it intersects.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the frontier be reached. On the other hand, this *can* happen. It happens, for example, when $d\gamma_{ij}/d\$ = \text{constant} \times \gamma_{ij}$, which obviously satisfies (6.7). In this case the geodesic motion is one of pure dilation, and the square of the distance to the frontier is, apart from an unimportant factor, simply the volume of 3-space.

(4) In the case of a pure dilation it is obvious that a geometrical singularity (zero volume) *does* occur at the frontier. That geometrical singularities must also occur in many other cases as well follows from the readily verified relation

$$d\ {}^{(3)}R/d\$ = -\ {}^{(3)}R^{ij}d\gamma_{ij}/d\$, \tag{6.9}$$

which holds as long as condition (6.7) is satisfied. In Appendix A it is shown that most geodesics (i.e., all but a set of measure zero) strike the frontier at points where some of the γ_{ij} (and hence some of the $d\gamma_{ij}/d\$$) become infinite, even though γ itself vanishes. Except in special cases, therefore, expression (6.9) will acquire singularities at the frontier.

With these mathematical preliminaries in mind let us now have a look at quantum dynamics. It is helpful to begin by analyzing Eq. (5.5) in the WKB approximation, so as to make the maximum possible use of classical ideas. We write

$$\Psi[{}^{(3)}\mathcal{G}] = \mathfrak{A} \exp(i\mathfrak{W}), \tag{6.10}$$

where $\mathfrak{A}$ and $\mathfrak{W}$ are assumed to be real functionals satisfying (roughly) the restriction

$$|\delta\mathfrak{A}/\delta\gamma_{ij}| \ll |\mathfrak{A}\delta\mathfrak{W}/\delta\gamma_{ij}|. \tag{6.11}$$

The phase then satisfies the Hamilton-Jacobi equation[32]

$$G_{ijkl}\frac{\delta\mathfrak{W}}{\delta\gamma_{ij}}\frac{\delta\mathfrak{W}}{\delta\gamma_{kl}} = \gamma^{1/2}\ {}^{(3)}R, \tag{6.12}$$

while the amplitude satisfies the conservation law

$$\delta(\mathfrak{A}^2 G_{ijkl}\delta\mathfrak{W}/\delta\gamma_{kl})/\delta\gamma_{ij} = 0. \tag{6.13}$$

In addition, the $\mathfrak{X}$ constraints impose the restrictions

$$\left(\frac{\delta\mathfrak{W}}{\delta\gamma_{ij}}\right)_{.j} = 0, \quad \left(\frac{\delta\mathfrak{A}}{\delta\gamma_{ij}}\right)_{.j} = 0. \tag{6.14}$$

Each solution of the Hamilton-Jacobi equation (6.12) determines a family of solutions of the classical field equations (i.e., a family of Ricci-flat 4-geometries) having the following property: For every 3-geometry there exists one and only one member of the family which has the 3-geometry as a spacelike hypersection, i.e., for which the 3-geometry is to be found among the infinity of spacelike hypersections which the member admits. Once $\mathfrak{W}$ is given, each 3-geometry determines a unique 4-geometry. The 4-geometry may be computed by making the identification $\pi^{ij} = \delta\mathfrak{W}/\delta\gamma_{ij}$ and integrating the equation

$$\partial\gamma_{ij}/\partial x^0 = 2\alpha G_{ijkl}\delta\mathfrak{W}/\delta\gamma_{kl} + \beta_{i.j} + \beta_{j.i}, \tag{6.15}$$

which follows from (2.5) and (3.3).[33] The quantities α and β_i are, as always, completely arbitrary, at least in sufficiently small finite regions. (Some global restrictions will generally exist.)

It is not hard to verify that (6.15) does indeed yield a solution of the classical field equations for each initial 3-geometry. One simply differentiates (6.15) with respect to x^0 and replaces $\delta\mathfrak{W}/\delta\gamma_{kl}$ by its expression in terms of α, $\beta_{i.j}$, and $\partial\gamma_{ij}/\partial x^0$. One finds

$$\gamma_{ij,00} = (\ln\alpha)_{,0}(\gamma_{ij,0} - \beta_{i.j} - \beta_{j.i}) + \frac{\partial G_{ijkl}}{\partial\gamma_{mn}} \times G^{klrs}\gamma_{mn,0}(\gamma_{rs,0} - 2\beta_{r.s}) + 4\alpha G_{ijkl}\int \frac{\delta^2}{\delta\gamma_{kl}\delta\gamma_{m'n'}} \times \left(\alpha' G_{m'n'r's'}\frac{\delta\mathfrak{W}}{\delta\gamma_{r's'}} + \beta_{m'.n'}\right) d^3x'. \tag{6.16}$$

The integration which appears in the last term of this equation may be performed with the aid of the identities

$$G_{m'n'r's'}\frac{\delta^2\mathfrak{W}}{\delta\gamma_{kl}\delta\gamma_{m'n'}}\frac{\delta\mathfrak{W}}{\delta\gamma_{r's'}} = -\frac{1}{2}\left[\frac{\partial G_{mnrs}}{\partial\gamma_{kl}}\frac{\delta\mathfrak{W}}{\delta\gamma_{mn}}\frac{\delta\mathfrak{W}}{\delta\gamma_{rs}} + \gamma^{1/2}({}^{(3)}R^{kl} - \gamma^{kl}\ {}^{(3)}R)\right] \times \delta(x,x') + \tfrac{1}{2}\gamma'^{1/2}\gamma^{m'n'}\gamma^{r's'} \times (\delta_{m'r'}{}^{kl}{}_{.n's'} - \delta_{m\,n'}{}^{kl}{}_{.r's'}), \tag{6.17}$$

$$\left(\frac{\delta^2\mathfrak{W}}{\delta\gamma_{kl}\delta\gamma_{m'n'}}\right)_{.n'} = \frac{\delta\mathfrak{W}}{\delta\gamma_{n'r'}}(\tfrac{1}{2}\delta_{n'r'}{}^{kl}{}_{.}{}^{m'} - \delta^{m'}{}_{n'}{}^{kl}{}_{.r'}), \tag{6.18}$$

which are obtained by functionally differentiating Eqs. (6.12) and (6.14) and making use of (4.25). The result is a set of six local-field equations which, together with (6.12) and (6.14) re-expressed in terms of $\alpha, \beta_{i.j}, \gamma_{ij,0}$, are equivalent to the ten Einstein empty-space equations.

Let us now make the simplifying assumptions $\alpha_{,i} = 0$, $\beta_i = 0$. We then have

$$G^{ijkl}\gamma_{ij,0}\gamma_{kl,0} = 4\alpha^2\gamma^{1/2}\ {}^{(3)}R. \tag{6.19}$$

Let us also assume that the integral

$$I \equiv \int \gamma^{1/2}\ {}^{(3)}R d^3x \tag{6.20}$$

[32] The Hamilton-Jacobi equation for general relativity appears to have been first written down by A. Peres, Nuovo Cimento 26, 53 (1962).

[33] The inverse problem of constructing the $\mathfrak{W}$ which corresponds to a given family of solutions of the classical field equations has been analyzed in detail by U. H. Gerlach (to be published). The author is indebted to Gerlach for the opportunity of studying this analysis in manuscript prior to publication.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

(extended over the whole of 3-space) is nonvanishing. We may then choose

$$\alpha=\tfrac{1}{2}|I|^{-1/2}, \tag{6.21}$$

which permits x^0 to be identified with the arc length $\$$ in the manifold $M^{\infty 3}$ and permits the first of Eqs. (6.14) to be re-expressed in the form

$$(G^{ijkl}d\gamma_{kl}/d\$)_{,j}=0, \tag{6.22}$$

which is identical with (6.7), showing that x^0 is in fact also the arc length in $\mathfrak{M}$. Finally, Eq. (6.16) takes the form

$$\frac{d^2\gamma_{ij}}{d\$^2}-\frac{d\gamma_{ik}}{d\$}\gamma_{kl}\frac{d\gamma_{ij}}{d\$}+\frac{1}{2}\frac{d\gamma_{ij}}{d\$}\gamma_{kl}\frac{d\gamma_{kl}}{d\$}+\tfrac{1}{8}\gamma^{-1/2}\gamma_{ij}G^{mnrs}\frac{d\gamma_{mn}}{d\$}\times\frac{d\gamma_{rs}}{d\$}=\frac{d\ln\alpha}{d\$}\frac{d\gamma_{ij}}{d\$}-2\alpha^2({}^{(3)}R_{ij}-\tfrac{1}{4}\gamma_{ij}\,{}^{(3)}R). \tag{6.23}$$

If the right-hand side of Eq. (6.23) were zero, the sequence of 3-geometries (as $\$$ varies) would trace out a geodesic in $\mathfrak{M}$. The right-hand term may therefore be regarded as a "force" term which caused the actual "trajectory" of 3-space to deviate from a geodesic. The following important questions arise: Is this "force term" powerful enough to keep the trajectory from striking the frontier? If not, what does arrival at the frontier mean physically?

Before giving answers to these questions, let us first take a crude over-all look at some of the simple implications of Eq. (6.23). It is not difficult to verify that if this equation is multiplied by $\frac{1}{2}\gamma^{1/2}\gamma^{ij}$ and the result is integrated over 3-space, the following equation is obtained:

$$\frac{d^2V}{d\$^2}-\frac{d\ln\alpha}{d\$}\frac{dV}{d\$}=-\tfrac{1}{4}, \quad V\equiv\int\gamma^{1/2}d^3x, \tag{6.24}$$

or equivalently,

$$\frac{d^2V}{d\tau^2}=-I, \tag{6.25}$$

where τ is the "proper time":

$$d\tau/d\$=\alpha. \tag{6.26}$$

From this it follows that the curve of V as a function of τ is concave downward whenever I is positive. Under these circumstances an expanding world tends to "slow down" while a contracting world tends to accelerate towards collapse.

A case for which I is positive is that in which 3-space has the geometry of a 3-sphere. The geometry cannot, however, remain spherical more than instantaneously, since the right-hand side of Eq. (6.19) is then everywhere positive, which requires the vector $d\gamma_{ij}/d\$$ to be "spacelike" in M for all x, thus ruling out the possibility of a pure dilation. The derivative $d\gamma_{ij}/d\$$ must contain shearing components corresponding to the presence of the gravitational radiation which is, in fact, needed in order to "close up" the universe. On the other hand, the 3-geometry may still be spherical in a coarse-grained sense. That is, although the sign of $\gamma^{1/2}\,{}^{(3)}R$ may fluctuate at a fine-grained level due to the presence of gravitational waves, its mean value may approximate that of a 3-sphere. In this case Eq. (6.25) takes the approximate form

$$\frac{d^2V}{d\tau^2}\approx-6(4\pi^4V)^{1/3}, \tag{6.27}$$

leading to a total lifetime of the universe given by[34]

$$T\approx\tfrac{3}{8}(2\pi^2)^{-1/3}\int_0^{V_{\max}}\frac{dV}{(V_{\max}^{4/3}-V^{4/3})^{1/2}}=\sqrt{2}\,\mathrm{cn}^{-1}(0|\tfrac{1}{2})R_{\max}=2.62R_{\max}, \tag{6.28}$$

where $R_{\max}$ is the radius of maximum expansion. Here it is clear that the "force term" in Eq. (6.23) does not prevent the 3-geometry from striking the frontier.

In the general case there are two factors which govern the trajectory of 3-space. Firstly, the condition $\alpha_{,i}=0$, which has been adopted in the above discussion, is known to be a poor one for keeping the hypersurfaces, $x^0=$ constant, smooth. When these hypersurfaces are sandwiched together, as here, with spatially uniform intervals, they often quickly develop geometrical singularities which have nothing to do with the geometry of space-time.[35] Such singularities can usually be avoided simply by relaxing the condition $\alpha_{,i}=0$. However—and this is the second factor—it is now known from the work of Avez,[36] Penrose,[37] Hawking,[38] and Geroch[39] that a nontrivial *singularity in space-time* "almost always" occurs at some point in the history of any physically interesting universe. At such a point abandonment of the condition $\alpha_{,i}=0$ is of no use. 3-space will acquire a geometrical singularity anwyay. Thus, if the initial hypersurface is sufficiently close to the point of onset of a change in 3-space topology, or if a so-called "trapped 2-surface" [37] is on the point of being born within it, then it will develop a geometrical

[34] This is to be compared with $T=2R_{\max}$ for a Friedmann universe filled with radiation treated as an ideal gas. Note that it is not possible to use expression (6.28) as an upper bound on the lifetime of the universe. Although it is easy to show that it is the spherical geometry which, for fixed V, makes I stationary (i.e., independent of small variations in the metric), this stationary point is neither a maximum nor a minimum, and hence it is not possible to assert that $I\geq 6(4\pi^4V)^{1/3}$.

[35] The phenomenon occurs already in a flat space-time. It is not possible to construct a family of uniformly spaced *curved* spacelike hypersurfaces in Minkowski space without the members of the family developing a geometrical singularity either in the past or in the future. The singularity always develops in the *convex* direction, contrary to the situation in a Euclidean manifold.

[36] A. Avez, Ann. Inst. Fourier (Grenoble) 13, 105 (1963).

[37] R. Penrose, Phys. Rev. Letters 14, 57 (1965).

[38] S. W. Hawking, Phys. Rev. Letters 17, 444 (1966).

[39] R. P. Geroch, Phys. Rev. Letters 17, 445 (1966).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

singularity which has nothing to do with the maintenance of the condition $\alpha_{,i}=0$. In this case the force term in (6.32) is again powerless to prevent the trajectory of 3-space from striking the frontier; indeed it may hasten the impact.

The occurrence of a singularity in space-time itself is known as *gravitational collapse*. Gravitational collapse may involve the whole of 3-space, as when the volume of the universe goes to zero, or it may involve only a small part of it (e.g., a collapsing superstar). It seems extremely likely that the almost universal inevitability of gravitational collapse is closely connected to the existence of the frontier in $\mathfrak{M}$. However, the establishment of this connection in rigorous terms is a major problem which remains unsolved. The existence proofs of Refs. 36–39 give no indication of the precise physical nature of the collapse singularity except for the statement that the normal causal properties of space-time break down there. This alone, of course, is enough to guarantee that the singularity represents a real barrier beyond which it is impossible to extend the solution of Einstein's equations. It means that in certain regions of the universe (or in the universe as a whole) time for the classical physicist, ultimately comes to an end beyond which he can make no further predictions.

The question now arises whether the classical collapse barrier, which we shall denote by $\mathfrak{B}$, is also a barrier for the solutions of the quantum equation (5.5). That the answer is not obvious may be seen as follows. Consider first a point ${}^{(3)}\mathcal{G}$ in $\mathfrak{M}$ which is not on $\mathfrak{B}$. If ${}^{(3)}\mathcal{G}$ has a singularity this must be due to the hypersurface $x^0=$ constant being chosen poorly. At the singular point in 3-space both the right and left sides of Eq. (6.12) will diverge. Correspondingly the two terms inside the parentheses of Eq. (5.5) will each contribute a divergence at this point. The two divergences will, however, cancel so that Eq. (5.5) is still satisfied. Consider now a point on $\mathfrak{B}$. Here something special happens which causes Eq. (6.12) to break down. However, it does not automatically follow that Eq. (5.5) likewise breaks down, for there exist possibilities for treating Eq. (5.5) which have no counterparts in the classical theory. For example, we note that if $\mathcal{W}$ is a solution of Eq. (6.12) then so is $-\mathcal{W}$. Moreover, the addition of an arbitrary constant to $\mathcal{W}$ leaves Eq. (6.12) unaffected. Let this constant be adjusted so that $\mathcal{W}$ vanishes at the point on $\mathfrak{B}$ in question, and choose for the WKB form of the solution of (5.5), the *superposition*

$$\Psi=\mathfrak{A}[\exp(i\mathcal{W})-\exp(-i\mathcal{W})]. \tag{6.29}$$

Then Ψ itself vanishes at the barrier, and this might conceivably alleviate the singularity in Eq. (5.5) which would otherwise occur, and permit an extension of Ψ beyond the barrier.

If one is looking for an example on which to practice hand-waving arguments he might consider a situation in which 3-space is about to undergo a change in topology. It can be shown that a change of topology requires (a) the development of a geometrical singularity in 3-space and (b) a breakdown in the causal structure (e.g., hyperbolic signature) of space-time at the onset of the singularity. Therefore topological transitions cannot be handled classically. However, since the singularity in ${}^{(3)}\mathcal{G}$ need occur at only a single point of 3-space it may develop in such a way that the corresponding singularities of Eq. (5.5) all cancel. We are careful, of course, not to say that the singularities *will* cancel. No one really knows whether topological transitions can be handled quantum mechanically.

Although the classical and quantum barriers may not be identical, and although each may depend to some extent on the particular solution of the Hamilton-Jacobi equation (6.12), or of the "wave equation" (5.5), under consideration at the moment, it seems very probable that there exists an irreducible *core* which is common to all barriers. We have suggested that it may be possible to continue Ψ past 3-geometries which contain isolated singularities. However, it is extremely difficult to imagine how such a continuation could be performed beyond a 3-geometry which has a dense set of singularities, or which is singular at *all* of its points, e.g., a 3-space of zero volume. It is therefore likely that the following set theoretical inequality holds:

$$\text{orb}\prod_{\mathbf{x}} F(\mathbf{x})\subseteq\mathfrak{B}_Q\subseteq\mathfrak{B}, \tag{6.30}$$

where $\mathfrak{B}_Q$ denotes the quantum barrier. In the remainder of the paper we shall assume that this inequality does hold.

The fact that $\mathfrak{B}$ is not the empty set is an embarrassment to the classical physicist, for it means that his theory breaks down. The fact that $\mathfrak{B}_Q$ is not the empty set, however, is not necessarily embarrassing to the quantum physicist, for he may be able to dispose of it by simply imposing, on the state functional, the following condition:

$$\Psi[{}^{(3)}\mathcal{G}]=0 \quad \text{for all } {}^{(3)}\mathcal{G} \text{ on } \mathfrak{B}_Q. \tag{6.31}$$

Provided it does not turn out to be ultimately inconsistent, this condition, which is already suggested by (6.29), yields two important results. Firstly, it makes the probability amplitude for catastrophic 3-geometries vanish, and hence gets the physicist out of his classical collapse predicament. Secondly, it may permit the Cauchy problem for the "wave equation" (5.5) to be handled in a manner very similar to that of the ordinary Schrödinger equation. Thus let Σ be a hypersurface like that which appears in Eq. (5.19). Since the dimensionality of $\prod_{\mathbf{x}}F(\mathbf{x})$ (the orbit of which forms the "core" of $\mathfrak{B}_Q$) is the same as that of Σ, namely $5\times\infty^3$, it would appear that the specification of Ψ on Σ, together with the boundary condition (6.31), is equivalent to its specification on two hypersurfaces and hence suffices to determine $\Psi[{}^{(3)}\mathcal{G}]$ completely for all ${}^{(3)}\mathcal{G}$.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

If this heuristic argument [based on the analogy of Eq. (5.5) to the Klein-Gordon equation] is indeed valid, then it is not necessary to specify also the normal derivatives of Ψ on Σ, despite the fact that Eq. (5.5) is of the second differential order.[40]

The only obvious difficulty with condition (6.31) is that it makes the presence of "negative frequency" components in Ψ unavoidable (see the discussion at the end of Sec. 5) and hence leaves one very unclear as to how to use Eq. (5.19) to define inner products and at the same time maintain positive definiteness of probability. In the following sections we shall show how this difficulty can be resolved in a special case.

7. THE QUANTIZED FRIEDMANN UNIVERSE

The simplest classical model which exhibits the collapse phenomenon is the Friedmann universe. If the Friedmann universe is assumed to be closed it must be filled either with gravitational radiation or with some other form of energy. It is not difficult to show that when other forms of energy are present in addition to gravity, the Hamiltonian constraint condition (4.3) is replaced by

$$(\mathfrak{C}+\mathfrak{M})\Psi=0, \tag{7.1}$$

where $\mathfrak{M}$ is the Hamiltonian of the system (or systems) giving rise to the additional energy. In order to avoid having to deal with entities as complicated as gravitons, with their spin and orbital states and their mutual interactions, we shall make use of such additional energy in the form of noninteracting material particles "at rest". The Friedmann universe is obtained by distributing these particles uniformly throughout a 3-sphere and "freezing out" all the degrees of freedom of the gravitational field save one, namely, that which corresponds to the time-varying spherical radius R.

If $\gamma^0{}_{ij}$ denotes the metric (in some coordinate system) of a 3-sphere of unit radius, then the 4-metric of the Friedmann world may be written in the form

$$(g_{\mu\nu})=\begin{pmatrix}-\alpha^2 & 0\\ 0 & \gamma_{ij}\end{pmatrix}, \quad \gamma_{ij}=R^2\gamma^0{}_{ij}, \tag{7.2}$$

where α and R depend only on x^0. Substituting this into (2.6), integrating over the volume $2\pi^2R^3$ of the Friedmann universe, and remembering that ${}^{(3)}R$ for a 3-sphere is $6/R^2$, we obtain for the effective Lagrangian of the gravitational field

$$L=12\pi^2[-\alpha^{-1}R(R_{,0})^2+\alpha R]. \tag{7.3}$$

As for the material particles (dust) which fill the universe, we shall, for reasons which will become clear as the analysis proceeds, endow them with internal dynamical degrees of freedom which may be described by canonical coordinates q^i and Lagrangians of the form $l(q,\dot{q})$, the dot denoting differentiation with respect to *proper* time:

$$\dot{q}^i=\alpha^{-1}q^i{}_{,0}. \tag{7.4}$$

Just as we have done for the gravitational field, however, we shall "freeze out" all the internal degrees of freedom save a small number by requiring all the particles to be identical and to be in coherent identical states (i.e., "in step"). Under these conditions the effective particle Lagrangian becomes

$$\mathrm{L}=\alpha N l(q,\alpha^{-1}q_{,0}), \tag{7.5}$$

where N is the total number of the particles in the universe.

Adding (7.3) and (7.5) to obtain the total Lagrangian we see that once again we have the primary constraint

$$\pi=\partial(L+\mathrm{L})/\partial\alpha_{,0}=0. \tag{7.6}$$

The wave function of the quantized Friedmann universe therefore cannot depend on α.

The total Hamiltonian becomes

$$\begin{aligned} H+\mathrm{H}&=\pi\alpha_{,0}+\Pi R_{,0}+P_iq^i{}_{,0}-L-\mathrm{L}\\ &=\pi\alpha_{,0}+\alpha(\mathfrak{C}+\mathfrak{M}), \end{aligned} \tag{7.7}$$

where

$$\Pi=\partial L/\partial R_{,0}=-24\pi^2\alpha^{-1}RR_{,0}, \tag{7.8}$$

$$P_i=\partial\mathrm{L}/\partial q^i{}_{,0}=Np_i, \quad p_i=\partial l/\partial\dot{q}^i, \tag{7.9}$$

$$\mathfrak{C}\equiv-\Pi^2/48\pi^2R-12\pi^2R, \tag{7.10}$$

$$\mathfrak{M}\equiv Nm, \quad m\equiv p_i\dot{q}^i-l. \tag{7.11}$$

The symbol m is here used to denote the internal Hamiltonian of the particles because the Hamiltonian is, in fact, the rest mass, provided the arbitrary zero point of the Lagrangian l has been properly chosen. We note that the "kinetic energy" term in the gravitational Hamiltonian (7.10) has the opposite sign (i.e., negative) from that of conventional Hamiltonians. This is because the only motion permitted to a Friedmann universe is one of pure dilation, and hence the coordinate R is "timelike."

The condition $\pi_{,0}=0$ leads immediately to the dynamical constraint $\mathfrak{C}+\mathfrak{M}=0$ which, in the quantum theory, takes the form (7.1). In the R representation this becomes

$$\left(\frac{1}{48\pi^2}R^{-1/4}\frac{\partial}{\partial R}R^{-1/2}\frac{\partial}{\partial R}R^{-1/4}-12\pi^2R+Nm\right)\Psi=0, \tag{7.12}$$

where factors have been ordered in such a way that the first term inside the parentheses becomes a one-dimensional Laplace-Beltrami operator,[41] and m is now the particle mass operator. Equation (7.12), which is

[40] Alternative and more detailed heuristic arguments leading to the same conclusion have been given by H. Leutwyler, University of Bern report, 1965 (unpublished).

[41] Here a definite ordering must be chosen. Since the number of degrees of freedom is now finite the ordering question cannot be treated as a problem in interpreting formally divergent symbols.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the analog of (5.5), must account completely for all the physical properties of the Friedmann universe.

It is by no means obvious how the familiar properties of the Friedmann world are to be extracted, in the classical limit, from Eq. (7.12), nor is it obvious what significance is to be attached to Ψ in the purely quantum domain. Difficulties of this type are not new to physics. A similar problem faced Schrödinger when he first wrote down the equation of the hydrogen atom. In his case there was a period of intense discussion, largely guided by Bohr, which ultimately led most physicists, with only a few dissenters of whom Einstein was the champion, to accept what has come to be known as the "Copenhagen view." The Copenhagen view depends on the assumed *a priori* existence of a classical level to which all questions of observation may utlimately be referred. Here, however, the whole universe is the object of inspection; there is no classical vantage point, and hence the interpretation question must be re-argued from the beginning. While we do not wish to stress this point unduly, since, after all, the Friedmann model ignores the vast complexities of the real universe, it is nevertheless clear that the quantum theory of space-time must ultimately force a deviation from the traditional Copenhagen doctrine.

Leaving aside these questions for the moment, let us note some of the simple mathematical properties of Eq. (7.12). If we carry out the point transformation

$$X\equiv R^{3/2},\quad \Phi\equiv(\partial R/\partial X)^{1/2}\Psi=(\tfrac{2}{3})^{1/2}R^{-1/4}\Psi, \quad (7.13)$$

Eq. (7.12) is converted to

$$-(3/64\pi^2)\partial^2\Phi/\partial X^2+12\pi^2X^{2/3}\Phi=Nm\Phi. \quad (7.14)$$

If the particles are in eigenstates of mass, so that m may be treated as a c number, and if the boundary condition

$$\Phi=0 \text{ at } X=0 \text{ or, equivalently, } \Psi=0 \text{ at } R=0 \quad (7.15)$$

analogous to (6.31) is imposed, then Eq. (7.14) becomes simply the Schrödinger equation of a particle of mass $32\pi^2/3$ moving at energy Nm in the one-dimensional potential

$$\begin{aligned} V&=\infty, & X&<0,\\ V&=12\pi^2X^{2/3}, & X&>0. \end{aligned} \quad (7.16)$$

Now unless the mass eigenvalue m happens to be such that Nm is one of the allowed eigenvalues of Eq. (7.14), the function Φ will not be normalizable but will behave in an exponential manner for large values of X. This is not necessarily bad if we insist on viewing R, and hence X, as a "time" coordinate, for it is usually impossible to require that a state function be normalizable with respect to *time*. Moreover, we may hesitate to allow the universe as a whole to determine the spectrum of masses which we can put into it, for in the classical theory the universe exerts no such control. However, several convincing arguments can be adduced which suggest that Φ must nonetheless be normalizable. The most important of these is that a closed Friedmann universe has, in the classical theory, a maximum radius of expansion. Hence if a correspondence principle is to exist, based on a transition to a classical limit, R must be effectively bounded from above. The existence of the classical turning point, as is well known, corresponds to the restriction to normalizable state functions.

For present purposes it suffices to determine the normalizable solutions of Eq. (7.12) in the WKB approximation. From the phase integral condition[42]

$$\begin{aligned} n+\tfrac{3}{4}&=-(2\pi)^{-1}\oint \Pi dR\\ &=24\pi\int_0^{R_{\max}}[R(R_{\max}-R)]^{1/2}dR, \end{aligned} \quad (7.17)$$

$$R_{\max}\equiv Nm/12\pi^2, \quad (7.18)$$

we obtain the "energy" spectrum

$$Nm=[48\pi^2(n+\tfrac{3}{4})]^{1/2},\quad n=0,1,2,\cdots. \quad (7.19)$$

Computation of the normalized state function itself involves only elementary integrals. Inside the turning point it is found to have the form

$$\begin{aligned} \Psi=(2/\pi R_{\max})^{1/2}&[(R_{\max}/R)-1]^{-1/4}\\ &\times\sin\{6\pi^2[(2R-R_{\max})(R(R_{\max}-R))^{1/2}\\ &+R_{\max}^2\sin^{-1}((R/R_{\max})^{1/2})]\}, \end{aligned} \quad (7.20)$$

while outside it falls off to negligible values at distances of the order of $R_{\max}^{-1/2}$ beyond $R_{\max}$.[43]

In realistic situations this function has an enormous number of nodes. For a Friedmann world approximating the actual universe one finds, very roughly,

$$n\sim 10^{120}, \quad (7.21)$$

and if all the degrees of freedom of the real world were taken into account the number would be vastly greater. However, despite the enormity of the quantum number, the function (7.20) does not provide a classical description of the universe, for it is a static function, composed of standing waves undergoing neither expansion nor contraction. The standing waves may, to be sure, be regarded as a superposition of waves "traveling" in opposite directions, those "traveling" in the direction of expansion (increasing R) corresponding, by virtue of the "timelike" character of R, to the "positive frequency" components mentioned at the end of Sec. 5, and those "traveling" in the direction of contraction corresponding to "negative frequency"

[42] Here $n+\frac{3}{4}$ is used in place of the usual $n+\frac{1}{2}$ because of the "hard wall" character of the potential (7.16) for $X<0$.

[43] Cf. Ref. 11, p. 462.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

components.[44] However, in order to make this "travel" apparent we need some other coordinate besides R.

It is at this point that the internal particle dynamics enter the picture. The collective internal motion permits the particle ensemble to be used as a clock. Classically the temporal behavior of R may be determined by means of the correlation which exists between R and the q^i. This correlation is described by the solutions of the Hamilton-Jacobi equation

$$\mathfrak{H}(R,\partial\mathcal{W}/\partial R)+\mathfrak{H}(q,\partial\mathcal{W}/\partial q)=0. \tag{7.22}$$

We may assume the particle Lagrangian l to be that of a multiply periodic system. The constants of integration of Eq. (7.22) are then conveniently taken to be the action variables J_i of the collective Lagrangian L, and since the equation is obviously separable we have solutions of the form

$$\mathcal{W}=-W(R,J)+\mathrm{W}(q,\mathrm{J})+\text{const.}, \tag{7.23}$$

where

$$J=-(2\pi)^{-1}\oint \Pi dR=J(\mathrm{J}), \quad \Pi=-\partial W/\partial R, \tag{7.24}$$

the integral being taken over a complete expansion-contraction cycle of the Friedmann universe. For these solutions Eq. (7.22) takes the separated form

$$\frac{1}{48\pi^2 R}\left(\frac{\partial W}{\partial R}\right)^2+12\pi^2 R=\mathfrak{H}(q,\partial\mathrm{W}/\partial q)=E(\mathrm{J}), \tag{7.25}$$

where $E(\mathrm{J})$ is a certain function of the J_i.

The q^i are obtained as functions of R and the J_i by integrating the simultaneous equations

$$dq^i/dR=\mathrm{V}^i(q,\mathrm{J})/V(R,J), \tag{7.26}$$

where

$$\mathrm{V}^i=(\partial\mathfrak{H}/\partial P_i)_{P=\partial\mathrm{W}/\partial q}, \tag{7.27}$$

$$V=(\partial\mathfrak{H}/\partial\Pi)_{\Pi=-\partial W/\partial R}=(24\pi^2 R)^{-1}\partial W/\partial R. \tag{7.28}$$

The integrals of Eqs. (7.26) are not hard to obtain. If Eq. (7.25) is differentiated with respect to J_i, one finds

$$V\frac{\partial^2 W}{\partial R\partial J}\frac{\partial J}{\partial \mathrm{J}_i}=\mathrm{V}^i\frac{\partial^2\mathrm{W}}{\partial q^i\partial \mathrm{J}_i}=\frac{\partial E}{\partial \mathrm{J}_i}, \tag{7.29}$$

which, together with (7.26), yields

$$\frac{\partial^2\mathrm{W}}{\partial \mathrm{J}_i\partial q^i}dq^i=\frac{\partial J}{\partial \mathrm{J}_i}\frac{\partial^2 W}{\partial J\partial R}dR, \tag{7.30}$$

whence

$$-\frac{\partial W}{\partial J}\frac{\partial J}{\partial \mathrm{J}_i}+\frac{\partial \mathrm{W}}{\partial \mathrm{J}_i}=\delta^i, \tag{7.31}$$

where the δ^i are "phase constants." Equations (7.31) may be solved algebraically to express the q's in terms of R and the constants of integration J_i, δ^i. The δ_i determine the relative phases of the simultaneous oscillatory motions, and the J_i determine the amplitudes.

In the quantum theory an analogous correlation between R and the q^i can be established provided the state function has the form of a *superposition* of solutions of (7.12) corresponding to different eigenvalues of m. It is well known that a multiply periodic system cannot be used as a clock if it is in an eigenstate of energy. The uncertainty principle requires many different energy levels to be represented in its wave function. Here, however, we run into a very special difficulty which is peculiar to the quantum theory of space-time. The values which m can assume are already determined by the quantization condition (7.19) quite independently of the form of the particle Lagrangian l. Hence, unless the operators $\mathfrak{H}$ and $-\mathfrak{H}$ have at least one eigenvalue in common, *the Hamiltonian constraint* (7.1) *will have no solutions at all.* Equation (7.1) is unlike an ordinary time-independent Schrödinger equation in that it picks out only a single eigenvalue of the operator $\mathfrak{H}+\mathfrak{H}$. Moreover, the latter operator, being the sum of two operators having spectra bounded respectively from above and from below, has itself a spectrum which stretches from $-\infty$ to ∞.

For purposes of the present discussion we must *assume* not only that $\mathfrak{H}+\mathfrak{H}$ has a zero eigenvalue but that this eigenvalue is highly *degenerate*. We shall postpone until Sec. 10 a discussion of what the actual state of affairs may be in the real universe. For the present we concentrate on mathematical developments.

We shall confine ourselves to the WKB approximation and look for solutions of Eq. (7.1) of the form [cf. Eq (6.11)]

$$\Psi=A\exp[i(-W+\mathrm{W})], \tag{7.32}$$

where A is a real amplitude satisfying (hopefully) the inequalities [cf. Eq. (6.12)]

$$\left|\frac{\partial A}{\partial R}\right|\ll\left|A\frac{\partial W}{\partial R}\right|, \quad \left|\frac{\partial A}{\partial q^i}\right|\ll\left|A\frac{\partial \mathrm{W}}{\partial q^i}\right|. \tag{7.33}$$

A differential equation for A may be obtained by substituting (7.32) into (7.1). One finds

$$[\mathfrak{H}(R,-i\partial/\partial R-\partial W/\partial R)+\mathfrak{H}(q,-i\partial/\partial q+\partial\mathrm{W}/\partial q)]A=0. \tag{7.34}$$

When the inequalities (7.33) are satisfied the "big" terms of (7.34) already add up to zero by virtue of the Hamiltonian-Jacobi equation (7.22). In order to obtain conditions on A we must include the smaller, "higher-order" terms, and for this purpose it is convenient to introduce a smooth real test function $\varphi(R,q)$ which vanishes outside a finite closed region in the R–q

[44] Owing to the negative character of the kinetic-energy term of the Hamiltonian (7.10), the directions of "travel" of the exponential components of a standing wave are opposite to the conventional ones.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

manifold.[45] We then subtract the equation

$$\int_0^\infty dR \int dq\ \varphi A[\mathfrak{IC}(R, -i\partial/\partial R - \partial W/\partial R) + \mathfrak{IC}(q, -i\partial/\partial q + \partial \mathbf{W}/\partial q)]A = 0 \quad (7.35)$$

from its complex conjugate and use the Hermiticity of $\mathfrak{IC}$ and $\mathfrak{JC}$ to obtain

$$\int_0^\infty dR \int dq\ A[\varphi, \mathfrak{IC}(R, -i\partial/\partial R - \partial W/\partial R) + \mathfrak{JC}(q, -i\partial/\partial q + \partial \mathbf{W}/\partial q)]A = 0. \quad (7.36)$$

When the inequalities (7.33) are satisfied, this becomes approximately

$$i\int_0^\infty dR \int dq[(\partial\varphi/\partial R)V + (\partial\varphi/\partial q^\nu)V^\nu]A^2 = 0, \quad (7.37)$$

which, by virtue of the arbitrariness of φ, implies (after an integration by parts)

$$\partial(A^2V)/\partial R + \partial(A^2V^i)/\partial q^i = 0. \quad (7.38)$$

This equation, which is the analog of (6.14), assures conservation of the "flux of probability" in the R-q manifold.

The general solution of Eq. (7.38) can be obtained by making use of the relations

$$V\partial\delta^i/\partial R + V^j\partial\delta^i/\partial q^j = 0, \quad (7.39)$$

$$\partial(V\partial^2 W/\partial R\partial J)/\partial R = 0, \quad (7.40)$$

$$\partial(V^iD)/\partial q^i = 0, \quad (7.41)$$

where

$$D \equiv \det(D_i{}^j), \quad D_i{}^j = \partial^2\mathbf{W}/\partial q^i\partial \mathbf{J}_j, \quad (7.42)$$

and where δ^i, in Eq. (7.39) is regarded not as a constant of integration but as a function of R and the q^i, defined by (7.31). It is easy to see that Eq. (7.39) follows from (7.29) and (7.31). Equation (7.40) is obtained by differentiating (7.29) with respect to R, while (7.41) is obtained by differentiating (7.29) with respect to q^k and multiplying by the matrix $D^{-1}{}_k{}^i$ inverse to (7.42). From these relations it follows immediately that

$$A^2 = (\partial^2 W/\partial R\partial J)DF(\delta), \quad (7.43)$$

where F is an arbitrary function of the δ^i.

Actually the form of F is not arbitrary, since there are other differential equations which the state function (7.32) must satisfy in addition to (7.12), namely, the eigenvalue equations

$$J(R, -i\partial/\partial R)\Psi = J\Psi, \quad (7.44)$$

$$\mathbf{J}_i(q, -i\partial/\partial q)\Psi = \mathbf{J}_i\Psi. \quad (7.45)$$

[45] It is aways to be understood that the R-q manifold is restricted to positive values of R (*excluding* zero).

The operators $J(R, -i\partial/\partial R)$ and $\mathbf{J}_i(q, -i\partial/\partial q)$ are obtained by solving the equations

$$\Pi = -\partial W/\partial R, \quad P_i = \partial \mathbf{W}/\partial q^i \quad (7.46)$$

for the J's in terms of Π, R, and the P's and q's, making the replacements $\Pi \to -i\partial/\partial R$, $P_i \to -i\partial/\partial q^i$, and carrying out appropriate Hermiticity symmetrizations. Now

$$J(R, -i\partial/\partial R)\Psi = \{\exp[i(-W+\mathbf{W})]\} \times J(R, -i\partial/\partial R - \partial W/\partial R)A, \quad (7.47)$$

$$\mathbf{J}_i(q, -i\partial/\partial q)\Psi = \{\exp[i(-W+\mathbf{W})]\} \times \mathbf{J}_i(q, -i\partial/\partial q + \partial \mathbf{W}/\partial q)A. \quad (7.48)$$

Because of the identities

$$J(R, -\partial W/\partial R) \equiv J, \quad \mathbf{J}_i(q, \partial W/\partial q) \equiv \mathbf{J}_i \quad (7.49)$$

the "big" terms of (7.47) and (7.48) already yield Eqs. (7.44) and (7.45). Hence the "higher-order" terms must vanish. With the aid of the inequalities (7.33) and a test function φ, as before, one easily finds that this implies

$$\partial(A^2\partial J/\partial\Pi)/\partial R = 0, \quad (7.50)$$

$$\partial(A^2\partial \mathbf{J}_i/\partial P_i)/\partial q^j = 0. \quad (7.51)$$

But $\partial J/\partial\Pi = -(\partial^2 W/\partial R\partial J)^{-1}$ and $\partial \mathbf{J}_i/\partial P_j = D^{-1}{}_i{}^j$. Hence, substituting (7.43) into (7.47) and (7.48), and making use of the identity

$$\partial(DD^{-1}{}_i{}^j)/\partial q^j = 0. \quad (7.52)$$

which can be shown to hold by virtue of the symmetry of $\partial D_k{}^l/\partial q^j$ in j and k, one finds that F must be a *constant*, independent of the δ_i.

In order to obtain normalizable solutions of (7.12) the J's must be quantized. In addition, the branch-point behavior of the functions W and $\mathbf{W}$ at the classical turning points must be taken into account, and superpositions of the form (7.32) corresponding to the different branches must be employed. These superpositions are the standard WKB solutions.

Suppose we freeze out all the collective particle degrees of freedom save one, and suppose this degree of freedom corresponds to a motion of libration in a smooth potential. Then we may distinguish two branches of the function $\mathbf{W}$, a branch $\mathbf{W}^+$ which increases with increasing q and a branch $\mathbf{W}^-$ which decreases with the increasing q. Similarly we may distinguish two branches, W^+ and W^-, of the function W. These branches are determined only up to arbitrary constants. The constants may be adjusted so that the WKB solutions take the form

$$\Psi_n = A_n[\exp(-iW_n{}^+) + \exp(-iW_n{}^-)] \times [\exp(i\mathbf{W}_n{}^+) + \exp(i\mathbf{W}_n{}^-)], \quad (7.53)$$

where

$$W_n{}^\pm \equiv W^\pm(R, n+\tfrac{3}{4}), \quad \mathbf{W}_n{}^\pm \equiv \mathbf{W}^\pm(q, n+\tfrac{1}{2}), \quad (7.54)$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

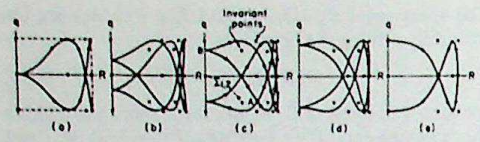


FIG. 1. Packet traces in the R-q plane. (Case $\Delta n=3$, $\Delta \mathrm{n}=2$.)

only those quantum numbers n, n being permitted for which

$$J=n+\tfrac{1}{4} \text{ when } \mathrm{J}=\mathrm{n}+\tfrac{1}{2}. \tag{7.55}$$

If the operator $\mathfrak{IC}+\mathfrak{JC}$ is to have a zero eigenvalue which is highly degenerate, it is clear from Eqs. (7.25) and (7.55) that the coherent internal dynamical behavior of the particles filling the universe must be precisely matched to that of the universe itself in such a way that the derivative $dJ/d\mathrm{J}$ is a *constant rational number* over a wide range of values of J. We shall write

$$dJ/d\mathrm{J}=\Delta n/\Delta \mathrm{n}, \tag{7.56}$$

where Δn and $\Delta \mathrm{n}$ are relatively prime integers. Δn and $\Delta \mathrm{n}$ are the spacings between adjacent permitted values of the quantum numbers n and n, respectively. An immediate consequence of Eq. (7.56) is that the angular frequencies (with respect to proper time) of the R and q motions are always (within the allowed range of J values) commensurable. These angular frequencies are given by

$$\omega \equiv dE/dJ=\omega d\mathrm{J}/dJ, \tag{7.57}$$

$$\omega \equiv dE/d\mathrm{J}, \tag{7.58}$$

and, in the allowed range, satisfy

$$\omega \Delta n=\omega \Delta \mathrm{n}. \tag{7.59}$$

Use of the angular frequencies permits Eqs. (7.29) to be re-expressed in the form

$$V\frac{\partial^2 W}{\partial R\partial J}=\omega, \quad \mathrm{V}\frac{\partial^2 W}{\partial q\partial \mathrm{J}}=\omega, \tag{7.60}$$

whence (7.43) becomes

$$A^2=F\omega\omega/V\mathrm{V}. \tag{7.61}$$

For later convenience we shall choose

$$F=\Delta n/2\pi\omega=\Delta \mathrm{n}/2\pi\omega. \tag{7.62}$$

The normalization constant A_n in expression (7.53) is then given by

$$A_n=(\omega_n \Delta n/2\pi|V_n \mathrm{V}_n|)^{1/2} \approx[(E_{n+\Delta n}-E_n)/2\pi|V_n \mathrm{V}_n|]^{1/2}, \tag{7.63}$$

the subscripts indicating that the quantized values of the quantities to which they are affixed are to be employed. No signs have been placed on the V's to correspond with the different branches of the W's, because motion in a potential is time-reversal invariant, and hence $|V^+|=|V^-|$, $|\mathrm{V}^+|=|\mathrm{V}^-|$.

8. A WAVE PACKET FOR THE UNIVERSE. THE CONCEPT OF TIME

We are now in a position to construct a state function exhibiting classical behavior. We do this by superposing many WKB solutions:

$$\Psi_a=\sum_n{}' a_n \Psi_n, \tag{8.1}$$

the prime indicating that the summation is to be carried out only over those quantum numbers which satisfy condition (7.55). If the a's are carefully chosen, Ψ_a will have the form of a "wave packet" which traces out a classical trajectory, namely, a generalized lissajous figure in the R-q plane. It is easy to see that Eq. (7.31) is just the condition for constructive interference, provided the symbols in the equation are understood to denote the "peak" values of the quantities to which they refer.

In view of the commensurability condition (7.59) the lissajous figure is necessarily closed and of finite length.[46] A typical set of packet traces is shown in Fig. 1 for the case $\Delta n=3$, $\Delta \mathrm{n}=2$. The action variable J is the same for each trace, but the phase constant δ varies from one to the other. The following facts may be inferred from the figures: Each trace lies within a rectangle having sides equal to the full amplitudes of the oscillations. Except in the degenerate cases depicted in Figs. 1(a) and 1(e), each trace divides the rectangle into $2\Delta n\Delta \mathrm{n}+\Delta n+\Delta \mathrm{n}+1$ disjoint regions. Although the size and shape of corresponding regions vary from figure to figure, each region contains an *invariant point* which is independent of δ. These points are shown in the figures.

The degenerate curves are those which have collapsed onto the invariant points. They are divided by the invariant points into a total of $4\Delta n\Delta \mathrm{n}$ segments, which will be called *invariant segments*. The invariant segments may be labeled in a systematic fashion, starting, say, from the "southwest" corner of the enclosing rectangle, by pairs of integers (r,r) satisfying $1\leq r\leq 2\Delta n$, $1\leq \mathrm{r}\leq 2\Delta \mathrm{n}$. When an invariant segment is used as a contour of integration (see Sec. 9) it will be denoted by the symbol $\Sigma_{r,\mathrm{r}}$. [See Fig. 1(c).] Each invariant segment corresponds to a lapse of proper time of amount $T/4\Delta n=\mathrm{T}/4\Delta \mathrm{n}$ where T and T are the oscillation periods.

[46] Degenerate forms of closure [see, for example, Figs 1(a) and 1(e)], in which the packet "moves" back and forth along the same curve, are to be understood as included in this statement. Also, if successive groups of a's are chosen to vanish in such a way that the *effective* spacing between adjacent quantum numbers becomes a multiple of Δn (or $\Delta \mathrm{n}$), then the packet trace will consist of m separately closed Lissajous figures superimposed upon one another. Such a trace must be understood as representing a *single* packet which consists of m disconnected parts. Although the following discussion can be extended to include such situations, they will, for simplicity, be excluded from consideration.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

A wave packet will be called *good* if its state function has negligible values throughout most of each region containing an invariant point. Except at intersection points or turning points only one of the branches of each of the functions W and $\mathbf{W}$ is involved in the constructive interference of the WKB functions at any one position along a good packet trace. Thus, for "motion" in a "northeasterly" direction the relevant branches are W^+, $\mathbf{W}^+$, provided we use the standard convention that time increases in the direction of increasing W and $\mathbf{W}$. Continuing around the compass we have W^+, $\mathbf{W}^-$ for SE; W^-, $\mathbf{W}^-$ for SW; and W^-, $\mathbf{W}^+$ for NW. The trace is thus divided into *branch segments* having definite quadrant orientations.

A given branch segment may be intersected by other branch segments which further subdivide it. The resulting pieces will be called *simple segments*. Each simple segment is intersected by precisely one invariant segment, and the two may therefore be labeled by the same integers. The branches involved in a given simple segment are W^+, $\mathbf{W}$, or W^-, $\mathbf{W}^-$ if $r+\mathbf{r}$ is odd and W^+, $\mathbf{W}^-$, or W^-, $\mathbf{W}^+$ if $r+\mathbf{r}$ is even, the choice depending on the direction of "motion." The direction in which the packet moves as time increases may be indicated by affixing arrows to the packet trace, as in Fig. 1(c). If the W's are adjusted so that $\delta=0$ corresponds to a degenerate trace, then the arrows are reversed by changing the sign of δ.

Proper time itself is defined by

$$\tau=\omega^{-1}\Theta\,,\quad \boldsymbol{\tau}=\boldsymbol{\omega}^{-1}\boldsymbol{\Theta}\,, \tag{8.2}$$

where the ω's are defined by Eqs. (7.57) and (7.58), and

$$\Theta=\partial W/\partial J\,,\quad \boldsymbol{\Theta}=\partial\mathbf{W}/\partial\mathbf{J}\,. \tag{8.3}$$

Classically the angle variables Θ and $\boldsymbol{\Theta}$ are canonically conjugate to $-J$ and $\mathbf{J}$, respectively, and hence τ and $\boldsymbol{\tau}$ are canonically conjugate to $\mathcal{H}$ and $\boldsymbol{\mathcal{H}}$, respectively. In the quantum theory this leads one to write the commutation relations

$$[\tau,\mathcal{H}]=i\,,\quad [\boldsymbol{\tau},\boldsymbol{\mathcal{H}}]=i\,. \tag{8.4}$$

It is important to remember, however, that the quantum τ's are *not Hermitian*. This follows not only from their periodic character, which arises from their dependence on the Θ's [Eqs. (8.2)], but also from the fact that their canonical conjugates, $\mathcal{H}$ and $\boldsymbol{\mathcal{H}}$, have discrete, "one-sided" spectra, bounded, respectively, from above and below. The usual eigenvector properties which hold for Hermitian operators therefore do not hold for the τ's, and we must distinguish between right and left eigenvectors.

Let us introduce the left eigenvectors $\langle\tau',\boldsymbol{\tau}'|$ which, in virtue of (8.4), may be chosen to satisfy

$$-i\frac{\partial}{\partial\tau'}\langle\tau',\boldsymbol{\tau}'|=\langle\tau',\boldsymbol{\tau}'|\mathcal{H},\quad -i\frac{\partial}{\partial\boldsymbol{\tau}'}\langle\tau',\boldsymbol{\tau}'|=\langle\tau',\boldsymbol{\tau}'|\boldsymbol{\mathcal{H}}. \tag{8.5}$$

We may also introduce the corresponding conjugate vectors, denoted by $|\tau',\boldsymbol{\tau}'\rangle$, which are right eigenvectors of the conjugate operators $\tau^\dagger$ and $\boldsymbol{\tau}^\dagger$. Now let $\mathcal{P}$ denote the projection operator into the physical subspace of allowed state vectors. Using Eqs. (8.5) and the Hamiltonian constraint (7.1), which may be rewritten in the form

$$(\mathcal{H}+\boldsymbol{\mathcal{H}})\mathcal{P}=\mathcal{P}(\mathcal{H}+\boldsymbol{\mathcal{H}})=0\,,\quad \mathcal{P}^2=\mathcal{P}\,, \tag{8.6}$$

it is easy to see that $\langle\tau,\boldsymbol{\tau}|\mathcal{P}$ depends only on the difference $\tau-\boldsymbol{\tau}$. This simple dependence may be recognized as a quantum consequence of the classical correlation

$$\tau-\boldsymbol{\tau}=-\omega^{-1}\delta\,, \tag{8.7}$$

which follows from (7.31) and (8.2).

The projection operator $\mathcal{P}$ is conveniently defined in terms of the eigenvectors $|n+\frac{1}{4},\ \mathbf{n}+\frac{1}{4}\rangle$ of the J's. Writing $|n+\frac{1}{4},\ \mathbf{n}+\frac{1}{4}\rangle\equiv|n\rangle$ whenever the quantum numbers are restricted as in (7.55), we have

$$\mathcal{P}=\textstyle\sum_n{}'|n\rangle\langle n|\,, \tag{8.8}$$

provided the normalization $\langle n|n'\rangle=\delta_{nn'}$ is assumed. In virtue of Eqs. (8.5) we may also write

$$\langle\tau,\boldsymbol{\tau}|n\rangle=(\omega_n\Delta n/2\pi)^{1/2}\exp[-iE_n(\tau-\boldsymbol{\tau})]. \tag{8.9}$$

The normalization here is chosen so as to maximize the orthogonality properties of the vectors $\langle\tau,\boldsymbol{\tau}|$ relative to the physical subspace. Noting that $\omega_n\Delta n$ gives the approximate spacing between adjacent permitted "energy" eigenvalues E_n, we have

$$\begin{aligned}\langle\tau,\boldsymbol{\tau}|\mathcal{P}|\tau',\boldsymbol{\tau}'\rangle&\\ \approx(2\pi)^{-1}\textstyle\sum_n{}'(\exp\{-iE_n[(\tau-\boldsymbol{\tau})-(\tau'-\boldsymbol{\tau}')]\})\Delta E_n&\\ =\delta((\tau-\boldsymbol{\tau})-(\tau'-\boldsymbol{\tau}'))\,.&\end{aligned} \tag{8.10}$$

If the "energy" spectrum were continuous and ranged from $-\infty$ to ∞ the function δ would be the Dirac δ. In reality it is a function which although divergent at the origin does not completely vanish elsewhere. Thus the eigenvectors $\langle\tau,\boldsymbol{\tau}|$ are only approximately orthonormal, a fact which stems from the lack of strict Hermiticity of the operators τ and $\boldsymbol{\tau}$.

Instead of working with the vectors $\langle\tau,\boldsymbol{\tau}|$ it is more interesting to work with $\langle\tau,q|$, $\langle R,\boldsymbol{\tau}|$, and $\langle R,q|$, which are defined in an obvious fashion. The normalization of $\langle R,q|$ may be fixed by setting

$$\langle R,q|n\rangle=\Psi_n\,, \tag{8.11}$$

where Ψ_n is the function having the WKB approximation (7.53), with A_n given by (7.63). In a similar manner the normalization of $\langle\tau,q|$ and $\langle R,\boldsymbol{\tau}|$ may be fixed by giving the WKB approximations of their inner products with $|n\rangle$. We shall choose

$$\begin{aligned}\langle\tau,q|n\rangle\equiv\Phi_n=(\omega_n\Delta n/2\pi|\,V_n\,|)^{1/2}&\\ \times e^{-iE_n\tau}[\exp(iW_n{}^+)+\exp(iW_n{}^-)]\,,&\end{aligned} \tag{8.12}$$

$$\begin{aligned}\langle R,\boldsymbol{\tau}|n\rangle=\Phi_n=(\omega_n\Delta n/2\pi|\,V_n\,|)^{1/2}[\exp(-iW_n{}^+)&\\ +\exp(-iW_n{}^-)]\exp(iE_n\boldsymbol{\tau})\,.&\end{aligned} \tag{8.13}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

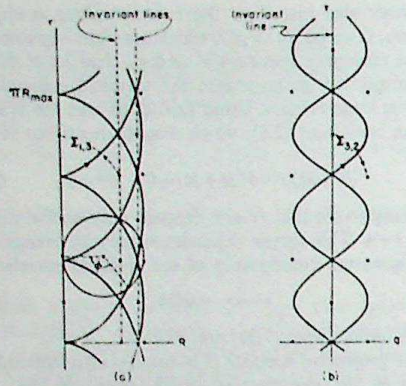


FIG. 2. Packet traces in the (a) R-τ and (b) τ-q planes. (Case $\Delta n=3$, $\Delta n=2$.)

Denoting by Φ and Φ arbitrary superpositions of the functions Φ_n and Φ_n, respectively, we may write the Schrödinger equations

$$i\partial\Phi/\partial\tau=\mathfrak{H}(q,-i\partial/\partial q)\Phi, \tag{8.14}$$

$$i\partial\Phi/\partial\tau=\mathfrak{H}(R,-i\partial/\partial R)\Phi. \tag{8.15}$$

When the function Ψ_a of (8.1) has the form of a wave packet so also have the corresponding functions Φ_a and Φ_a. The form of the packet trajectory in the case of the function Φ_a may be determined by noting that the condition for constructive interference, which establishes the correlation between R and τ, is

$$\partial W/\partial J=\omega\tau+\delta, \tag{8.16}$$

where δ is a phase constant and the other symbols denote the peak values of the quantities to which they refer.[47] Differentiation of Eq. (8.16) with respect to τ and use of Eqs. (7.28) and (7.60) yields

$$dR/d\tau=V=-(24\pi^2R)^{-1}\Pi$$
$$=R^{-1}[R(R_{max}-R)]^{1/2}. \tag{8.17}$$

The integration of this equation is most easily carried out with the aid of an angle ϕ defined by

$$d\tau/d\phi=R. \tag{8.18}$$

This gives

$$d\phi=[R(R_{max}-R)]^{-1/2}dR, \tag{8.19}$$

which yields the familiar cyclodial trajectory of the dust-filled Friedmann universe:

$$R=\tfrac{1}{2}R_{max}(1-\cos\phi), \tag{8.20}$$

$$\tau=\tfrac{1}{2}R_{max}(\phi-\sin\phi), \tag{8.21}$$

the constants of integration being chosen so that $R=0$, $\phi=0$ at $\tau=0$. We note that by virtue of the boundary condition (7.15) the packet rebounds repeatedly from the collapsed state until it ultimately loses its identity owing to spreading. Throughout the period of each rebound the width of the packet remains at all times finite, never suffering infinite compression. Transition through collapse thus becomes, in the quantum theory, a continuous process—something which cannot be achieved within the classical framework.

Figure 2 shows the curves traced out in the R-τ and τ-q planes by the packet of Fig. 1, and reveals a slight complication which was overlooked in the above simple analysis. The curves in the two planes appear to depict Δn and Δn distinct packets, respectively, rather than only a single packet. The extra "ghost packets" arise because the complete spectra of $\mathfrak{H}$ and $\mathfrak{H}$ are not made use of in the superposition (8.1). Only every Δnth level of $\mathfrak{H}$ and every Δnth level of $\mathfrak{H}$ occur. This means that τ and τ are determined modulo $T/\Delta n=\mathrm{T}/\Delta$n rather than modulo T and T, respectively, and a given packet must consequently appear "simultaneously" in several places in order to allow for the resulting proper-time ambiguity.[48]

The multiple traces intersect themselves along $\Delta n-1$ and Δn-1 phase-invariant lines, respectively. (See the indicated lines in the figures.) These lines divide the *branch segments* (which are defined just as for the R-q lissajous figures) into simple segments. Each simple segment is straddled by a unique pair of points at which maximum destructive interference occurs. The pairs of points may be connected by arcs intersecting the associated simple segments. These arcs will be denoted by $\Sigma_{r,t}$ and $\Sigma_{t,r}$ in the R-τ and τ-q planes, respectively. The pairs of suffixes r, t and t, r have the ranges $r=1, 2\cdots\Delta n$; r$=1, 2, \cdots\Delta$n; t, t$=\cdots-2, -1, 0, 1, 2, \cdots$, and may be used in an obvious manner to identify either the simple segments or their associated intersecting arcs. Examples of arcs and point pairs are shown in the figure.

With the introduction of the three wave functions Φ, Φ, and Ψ we now have at our disposal three distinct mathematical windows from which to view the Friedmann world. From one window the material content of the universe is seen as a clock for determining the dynamical behavior of the world geometry. From another it is the geometry which appears as a clock for determining the dynamical behavior of the material content. From the third the geometry and the material content appear on equal footing, each one correlated in a certain manner with the other.

It is the third window which is to be preferred as most accurately revealing the physics of the quantized Friedmann model. The variables τ and τ, because of their lack of Hermiticity, are not rigorously observable and hence cannot yield a measure of proper time which is valid under all circumstances. It is only with good

[47] Equation (8.16) also follows from (8.2), (8.3), and (8.7).

[48] This has also the consequence that $\langle R,\tau|\Phi|R',\tau\rangle$ and $\langle\tau,q|\Phi|\tau,q'\rangle$ do not have the form of simple δ functions, although they diverge at $R=R'$ and $q=q'$.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

wave packets that these variables are useful. But even with a good packet the description in terms of τ and $\mathfrak{r}$ is not perfect, as is revealed in a striking way by the fact that the wave packets Φ_a and Φ_a inevitably spread in "time," whereas the packet Ψ_a does not. It is for this reason that we may say that "time" is only a phenomenological concept, useful under certain circumstances.

It is worth remarking that it is not necessary to drag in the whole universe to argue for the phenomenological character of time. If the principle of general covariance is truly valid then the quantum mechanics of every-day usage, with its dependence on Schrödinger equations of the form (8.14) or (8.15), is only a phenomenological theory. For the only "time" which a covariant theory theory can admit is an intrinsic time defined by the contents of the universe itself. Any intrinsically defined time is necessarily non-Hermitian, which is equivalent to saying that there exists no clock, whether geometrical or material, which can yield a measure of time which is operationally valid under *all* circumstances, and hence there exists no operational method for determining the Schrödinger state function with arbitrarily high precision. This statement also follows directly from the uncertainty principle. Because every clock has a "one-sided" energy spectrum, its ultimate accuracy must necessarily be inversely proportional to its rest mass. When the whole universe is cast in the role of a clock, the concept of time can of course be made fantastically accurate (at least in principle) because of the enormity of the masses and quantum numbers involved. But as long as the universe is finite, a theoretical limit to the accuracy nevertheless remains.

9. THE INNER PRODUCT

We shall now use the results of the two preceding sections to show how the definition (5.19) for inner products can be rescued from the negative-probability disaster, at least in the case of the quantized Friedmann model. First we must derive the form which (5.19) takes in this model. Consider the following integral:

$$\int \varphi\{[\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\Psi_b]^*\Psi_a-\Psi_b^*\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\Psi_a\}dq,$$

where Ψ_a and Ψ_b are arbitrary complex functions of R and the q^i, and φ is a real test function. Because of the Hermiticity of $\mathfrak{H}\mathcal{C}$ this integral may be rewritten in the form

$$\int \Psi_b^*[\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q),\varphi]\Psi_a dq = -i\int \Psi_b^*\mathrm{V}^i(q,-i\partial/\partial q)\bullet(\partial\varphi/\partial q^i)\Psi_a dq, \quad (9.1)$$

where V^i is defined by (7.27), but with the replacement $P_i\rightarrow -i\partial/\partial q^i$ instead of $P_i=\partial \mathrm{W}/\partial q^i$, and where the dot in the right-hand integrand indicates that the factor $\partial\varphi/\partial q^i$ is to be inserted between noncommuting factors in the terms of V^i in such a way as to yield the commutator on the left. If now the differential operators occurring in V^i are peeled to the left and right, via integrations by parts, in such a manner that they no longer act on $\partial\varphi/\partial q^i$, then the integral takes the form

$$-i\int(\partial\varphi/\partial q^i)(\Psi_b^*\overleftrightarrow{\mathrm{V}}^i\Psi_a)dq = i\int\varphi\partial(\Psi_b^*\overleftrightarrow{\mathrm{V}}^i\Psi_a)/\partial q^i dq, \quad (9.2)$$

where $\overleftrightarrow{\mathrm{V}}^i$ denotes the result of the peeling process. Because of the arbitrariness of φ it follows that

$$[\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\Psi_b]^*\Psi_a-\Psi_b^*\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\Psi_a = \partial(\Psi_b^*\overleftrightarrow{\mathrm{V}}^i\Psi_a)/\partial q^i. \quad (9.3)$$

In a similar manner we find

$$[\mathfrak{H}\mathcal{C}(R,-i\partial/\partial R)\Psi_b]^*\Psi_a-\Psi_b^*\mathfrak{H}\mathcal{C}(R,-i\partial/\partial R)\Psi_a = \partial(\Psi_b^*\overleftrightarrow{\mathrm{V}}\Psi_a)/\partial R, \quad (9.4)$$

where in this case we can give an explicit form for $\overleftrightarrow{V}$:

$$\overleftrightarrow{V}=\frac{i}{48\pi^2}\left(R^{-3/4}\frac{\overrightarrow{\partial}}{\partial R}R^{-1/4}-R^{-1/4}\frac{\overleftarrow{\partial}}{\partial R}R^{-3/4}\right). \quad (9.5)$$

The analog of (5.19) is now obvious, namely,

$$(\Psi_b,\Psi_a)=\int_\Sigma(\Psi_b^*\overleftrightarrow{V}\Psi_a dq+\Psi_b^*\overleftrightarrow{\mathrm{V}}^i\Psi_a dR d\Sigma_i), \quad (9.6)$$

where Σ is an appropriate surface in the R-q manifold and $d\Sigma_i$ is the directed surface element of its projection into q space. From Eqs. (9.3) and (9.4) it follows that

$$\partial(\Psi_b^*\overleftrightarrow{\mathrm{V}}\Psi_a)/\partial R+\partial(\Psi_b^*\overleftrightarrow{\mathrm{V}}^i\Psi_a)/\partial q^i=0, \quad (9.7)$$

whenever Ψ_a and Ψ_b are physical state functions satisfying the Hamiltonian constraint (7.1). Therefore the integral (9.6) is independent of Σ provided the boundary of Σ remains in a region where Ψ_a and Ψ_b vanish.

When the coherent dust filling the Friedmann universe is restricted to only one degree of freedom the inner product (9.6) reduces to

$$(\Psi_b,\Psi_a)=\int_\Sigma(\Psi_b^*\overleftrightarrow{V}\Psi_a dq-\Psi_b^*\overleftrightarrow{\mathrm{V}}\Psi_a dR), \quad (9.8)$$

where Σ is an appropriate contour in the R-q plane. The key word here is "appropriate." In analogy with our previous treatment of the manifold $\mathfrak{M}$ of 3-geometrics in the general theory, we may view the R-q plane as endowed with a natural metric determined by the structure of the functions $\mathfrak{H}\mathcal{C}$ and $\mathfrak{H}\mathcal{C}$. With respect to this metric the coordinates R and q are "timelike" and "spacelike", respectively. If the Hamiltonian constraint

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

(7.1) were an ordinary wave equation we would naturally adopt for the contour Σ a "spacelike" line such as $R=$ constant. However, just as in the general theory, so also here, "wave" propagation is not restricted to timelike directions. Indeed, from the lissajous traces of Fig. 1, it is evident that the Friedmann universe not only executes "timelike" and "spacelike" motions with impartiality, but even turns around and "moves" backward with respect to the "time" coordinate. The distinction between "timelike" and "spacelike" clearly does not have the same pervasive significance here as it does in ordinary wave theories.

If we were actually to choose, for Σ, a line $R=$ constant, we would obtain the useless result $(\Psi_b,\Psi_a)=0$ for all Ψ_a and Ψ_b. This is because all physically admissible state functions have non-negligible values only in a finite domain of the R-q plane. Hence any line $R=$ constant can be deformed into a line along which Ψ_a and Ψ_b effectively vanish, without affecting the value of the ntegral (9.8) at all. The same is true if Σ is a "timelike" curve which starts at $R=0$ and goes out to infinity in the R-q plane. Since any normalizabile superposition of the functions (7.20) vanishes at $R=0$ at least as fast as $R^{7/4}$, such a curve can also be deformed into one along which Ψ_a and Ψ_b vanish, without affecting (9.8).

How then shall we choose Σ? The answer is to be found in the conservation laws (6.13), (7.38), and (9.7). From our analysis of the Lissajous traces of Fig. 1 it is evident that probability flows in a closed finite circuit in the R-q plane. Σ must therefore be a *finite* curve, chosen so as to intersect a unidirectional unit flux of probability of each of the two functions Ψ_a and Ψ_b. This means that Eq. (9.8) can be used to define inner products only when Ψ_a and Ψ_b both have the form of good packets. If they do not have this form or if they fall into the degenerate category depicted in Figs. 1(a) and 1(e), then some other representation must be employed. An analogous condition must hold in the general theory if Eq. (5.19) is to be valid. Whenever the condition is violated the usefulness of Wheeler's "metric representation" diminishes.

It is not difficult to show that Eq. (9.8) indeed yields an acceptable value for the inner product under the required conditions. The case in which the two packets do not overlap (except at intersections) may be disposed of at once; (Ψ_b, Ψ_a) then clearly vanishes. The only case which need concern us is that in which the two packets overlap, at least partially, throughout the entire length of their trajectories. The J values at their "peaks" then differ negligibly compared to the spread of values contained in their superpositions. As our initial contour we shall choose an invariant segment $\Sigma_{r,t}$ corresponding to the average of the peak J values. Since the packets are "good" we know that Ψ_a and Ψ_b vanish at its endpoints. Suppose $\Sigma_{r,t}$ intersects a NE-SW branch of the lissajous figure formed by the packet traces [e.g., the segment $\Sigma_{1,2}$ shown in Fig. 1(c)], and suppose the orientation of both packets is the same, say NE, at the point of intersection with $\Sigma_{r,t}$. Then it is only the W^+, W^+ branch of each packet which interferes constructively along $\Sigma_{r,t}$. Approximating Ψ_a and Ψ_b by their WKB forms, keeping only those parts which refer to the branch in question, and taking note of the inequalities (7.33), we have

$$(\Psi_b,\Psi_a)=\int_{\Sigma_{r,t}}(\Psi_b^*\vec{\nabla}\Psi_a dq-\Psi_b^*\vec{\nabla}\Psi_a)dR$$

$$=\sum_{n,n'}{}' b_{n'}^* a_n \int_{\Sigma_{r,t}}(\omega_{n'}\omega_n\Delta n'\Delta n/4\pi^2|V_{n'}V_nV_{n'}V_n|)^{1/2}\{[\exp(iW_{n'}{}^+)\vec{\nabla}\exp(-iW_n{}^+)]$$

$$\times\exp[-i(W_{n'}{}^+-W_n{}^+)]dq-\exp[i(W_{n'}{}^+-W_n{}^+)][\exp(-iW_{n'}{}^+)\vec{\nabla}\exp(iW_n{}^+)]dR\}\,, \quad (9.9)$$

where the b's are the coefficients of the expansion of Ψ_b:

$$\Psi_b=\Sigma_n{}'b_n\Psi_n\,. \quad (9.10)$$

Having dropped the parts of the WKB functions which refer to irrelevant branches, we may now extend the contour $\Sigma_{r,t}$ until its ends coincide with points at which maximum destructive interference (of the W^+, W^+ parts) occurs [e.g., the points A and B in Fig. 1(c)]. The contour then corresponds to a proper time lapse of $T/2\Delta n$ instead of $T/4\Delta n$, and spans just the right number of nodes so that the integral in (9.9) vanishes except when $n=n'$. Expression (9.9) accordingly reduces to

$$(\Psi_b,\Psi_a)=\sum_n{}' b_n^* a_n\int_\Sigma[(\omega_n\Delta n/2\pi|V_n|)dq$$

$$-(\omega_n\Delta n/2\pi|V_n|)dR]$$

$$=\sum_n{}' b_n^* a_n[(\omega_n T/4\pi)+(\omega_n T/4\pi)]\,, \quad (9.11)$$

the positive sign of the final bracketed factor being obtained by appropriately orienting the original contour $\Sigma_{r,t}$. The contour $\Sigma_{1,2}$ in Fig. 1(c) shows the correct orientation. If the direction of "motion" of the packets

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

is reversed then the orientation of the contour must be reversed.

For good packets the frequencies ω_n and ω_n remain sensibly constant and equal to the peak frequencies $2\pi/T$ and $2\pi/\mathbf{T}$, respectively, over the range of effective n values in the sum (9.11). Therefore we have

$$(\Psi_b,\Psi_a)=\sum_n{}' b_n{}^* a_n\,, \tag{9.12}$$

which is just the accepted definition. By virtue of the Σ invariance of expression (9.8) the contour may now be displaced to any location, including turning points where the WKB approximation breaks down. All that is required is that the contour cut each wave packet only once and that Ψ_a and Ψ_b vanish at its endpoints. We therefore have quite generally

$$\int_\Sigma (\Psi_b{}^*\vec{\nabla}\Psi_a dq-\Psi_b{}^*\vec{\nabla}\Psi_a dR)\approx\sum_n{}' b_n{}^* a_n\,, \tag{9.13}$$

the relation "$\approx$" tending toward "$=$" the more precisely defined the packets Ψ_a and Ψ_b become.

If, in the above derivation, the two packets had been oppositely oriented then one of the pairs of functions W^+, $\mathbf{W}^+$ in Eq. (9.9) would have had to be changed to W^-, $\mathbf{W}^-$, and the integral, with the extended contour, would have vanished even when $n=n'$. This, however, does not conflict with (9.12) since, in the case of oppositely oriented packets, the relative phases of a_n and b_n vary so rapidly with n that the inner product vanishes anyway.

An entirely similar analysis can be carried out in the R-τ and τ-q planes. Here the inner product integrals are given by

$$(\Phi_b,\Phi_a)=\int_{\Sigma_{\tau,\,t}} (\Phi_b{}^*\vec{\nabla}\Phi_a d\tau-\Phi_b{}^*\Phi_a dR)\,, \tag{9.14}$$

$$(\Phi_a,\Phi_b)=\int_{\Sigma_{t,\,\tau}} (\Phi_b{}^*\Phi_a dq-\Phi_b{}^*\vec{\nabla}\Phi_a d\tau)\,, \tag{9.15}$$

which reduce to the familair $\int\Phi_b{}^*\Phi_a dR$ and $\int\Phi_b{}^*\Phi_a dq$ when the contours are distorted to $\tau=$ constant and $\tau=$ constant, respectively. If the ranges of integration of the latter integrals are extended to include all permissible R and q values, the integrals must be divided by Δn and $\Delta\mathbf{n}$, respectively, because of the presence of the ghost packets.

The above analysis permits us to adopt a new viewpoint regarding the Cauchy problem for the Hamiltonian constraint. At the end of Sec. 6 it was conjectured that by virtue of the boundary condition (6.31) [(7.15) in the present context] the state function will be determined everywhere as soon as it is specified on a hypersurface. This is very easy to demonstrate in the present context, because of the separability of Eq. (7.12), which permits the eigenfunctions Ψ_n to be expressed in the product form

$$\Psi_n(R,q)=(2\pi\Delta n/\omega_n)^{1/2}X_n(R)\mathbf{X}_n(q)\,, \tag{9.16}$$

where X_n and $\mathbf{X}_n$ have the WKB approximations

$$X_n=(\omega_n/2\pi|V_n|)^{1/2}\times[\exp(-iW_n{}^+)+\exp(-iW_n{}^-)]\,, \tag{9.17}$$

$$\mathbf{X}_n=(\omega_n/2\pi|\mathbf{V}_n|)^{1/2}\times[\exp(i\mathbf{W}_n{}^+)+\exp(i\mathbf{W}_n{}^-)]\,, \tag{9.18}$$

and satisfy the orthonormality conditions

$$\int_0^\infty X_n{}^* X_{n'}dR=\delta_{nn'}\,,\quad \int\mathbf{X}_n{}^*\mathbf{X}_{n'}dq=\delta_{nn'}\,. \tag{9.19}$$

Thus, we may write

$$\Psi(R,q)=\sum_n{}' X_n(R)\mathbf{X}_n(q)\int\mathbf{X}_n{}^*(q')\times\Psi(R',q')dq'/X_n(R') \tag{9.20a}$$

$$=\sum_n{}' X_n(R)\mathbf{X}_n(q)\int_0^\infty X_n{}^*(R')\times\Psi(R',q')dR'/\mathbf{X}_n(q')\,, \tag{9.20b}$$

which express Ψ everywhere in terms of its values on the infinite contour $R=R'$ or on the infinite contour $q=q'$.

However, when the function Ψ has the form of a wave packet Ψ_a, it should be equally possible to determine it completely by knowing its value over a *finite* contour Σ which intersects the packet only once. That this is indeed the case follows from the fact that for a good packet the integrals

$$\int_\Sigma (\Psi_n{}^*\vec{\nabla}\Psi_a dq-\Psi_n{}^*\vec{\nabla}\Psi_a dR)\,, \tag{9.21}$$

for all n, may to a high degree of accuracy be replaced simply by

$$\int_\Sigma (\Psi_n{}^*|V_n|\Psi_a dq-\Psi_n{}^*|\mathbf{V}_n|\Psi_a dR)\,. \tag{9.22}$$

When the packet is good these integrals have nonnegligible values only for a restricted range of n values centered on the peak of the packet, over which $|V_n|$ varies slowly. Let the peak n values be determined by evaluating (9.22) for all n. Then let the contour be extended until its ends reach points of maximum destructive interference, as determined by the peak n value and the slope of the packet branch where it intersects Σ.[49] Suppose the slope is NE-SW, corresponding to the classical functions W^+, $\mathbf{W}^+$ or W^-, $\mathbf{W}^-$. Denote by

[49] It may be objected that in choosing Σ to intersect the packet along a definite branch we are assuming some preliminary knowledge about the approximate "location" of the packet. This preliminary knowledge, however, differs in no fundamental respect from the knowledge which we always have in other more familiar instances, e.g., that a given particle is "somewhere in the laboratory."

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Ψ_n^{++} and Ψ_n^{--} the parts of (the WKB approximation to) Ψ_n associated with these functions. Now compute the integrals

$$\int_{\Sigma'} (\Psi_n^{\pm\pm *} | V_n | \Psi_a dq - \Psi_n^{\pm\pm *} | V_n | \Psi_a dR), \quad (9.23)$$

where Σ' denotes the extended contour. Of these integrals, only those corresponding to the previously determined significant n values need be included, and of these, in turn, only those corresponding to a definite choice of signs (either $++$ or $--$) will have non-negligible values (corresponding to a definite packet orientation, which becomes thus determined). The values in question are just the amplitudes a_n of Eq. (8.1) and from these the entire state function can be constructed. This means that for a good packet the Cauchy data are not only the same as for the ordinary Schrödinger equation but are also effectively taken from a compact domain of "configuration space."

10. DISCUSSION AND SPECULATION

Perhaps the most impressive fact which emerges from a study of the quantum theory of gravity is that it is an extraordinarily economical theory. It gives one just exactly what is needed in order to analyze a particular physical situation, but not a bit more. Thus it will say nothing about time unless a clock to measure time is provided, and it will say nothing about geometry unless a device (either a material object, gravitational waves, or some other form of radiation) is introduced to tell when and where the geometry is to be measured.[50] In view of the strongly operational foundations of both the quantum theory and general relativity this is to be expected. When the two theories are united the result is an operational theory *par excellence*.[51]

The economy of quantum gravidynamics is also revealed in the manner in which the formalism determines its own interpretation. We have seen how the Hamiltonian constraint, in the case of a finite universe, forces us to abandon all use of externally imposed coordinates (in particular x^0) and to look instead for an internal description of the dynamics. We have seen how the metric structure of the manifold $\mathfrak{M}$, with its frontier of infinite curvature, suggests a natural boundary condition for the state functional, which may simplify the Cauchy data needed to specify a state. And finally, if it be permitted to extend the results of our study of the Friedmann model to the general case, we have learned how (and when) to use the inner-product definition (5.19), by recognizing that probability flows in closed circuits in $\mathfrak{M}$.

This "principle of self-determination," which permeates even classical general relativity, has been elevated to the rank of a universal principle by Everett,[52] who applies it to ordinary nonrelativistic quantum mechanics. As conventionally formulated quantum mechanics comes in two packages: (1) formalism and (2) interpretation based on the existence of a classical level. According to Everett, package 2 should be thrown away. Quantum mechanics is a theory which attempts to describe in mathematical language a situation in which chance is not a measure of our ignorance but is *absolute*. Naturally it cannot avoid introducing things like wave functions which undergo repeated fission, corresponding to the many possible outcomes of a given physical process. According to Everett, the wave function nonetheless provides a faithful representation of reality; it is the universe itself which splits.

To those who would immediately object that they do not feel themselves split, Everett replies that this only confirms the theory; they are not supposed to feel it. Everett allows into the theory only those elements which are in the formalism itself, namely, a Hilbert space, a Hamiltonian, and a Schrödinger equation for vectors in the Hilbert space. From these meager beginnings one can show, by standard arguments, that the wave function for a Hamiltonian which, in conventional language, would be described as that of a system coupled to an apparatus, evolves into a superposition of vectors representing the possible values of some system variable together with corresponding apparatus "readings." Moreover, if the "measurement" is repeated on a large number N of identically prepared systems, the final superposition consists of vectors representing various possible sets of N values for the system variable together with corresponding apparatus "memory sequences" which record these values. No interpretation of the mathematics is admitted up to this point; in particular no *a priori* interpretation is given to the coefficients in the final superposition.

Now let the coefficients in the final superposition in the case of a single system be denoted by c_n. Then the coefficients in the case of N systems will be products of c's. It can be shown[53] that if one removes from the final N-system superposition all those vectors which correspond to memory sequences in which the recorded values of the system variable fail to meet the standard

[50] For details on the quantum theory of measurement in general relativity see B. S. DeWitt, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962).

[51] A notable failure to recognize this fact is to be found in P. W. Bridgman, in *Albert Einstein, Philosopher Scientist*, edited by P. A. Schilpp (Tudor Publishing Company, New York, 1949). Bridgman's confusion, which is shared by others, stems from the fact that in traditional formulations of general relativity one speaks about things, such as curvilinear coordinates, which have no operationally defined reality. This confusion would have been eliminated had modern coordinate-independent formulations of differential geometry been available in 1916. Modern methods make it plain that coordinate systems are precisely what general relativity is *not* talking about. General relativity is concerned with those attributes of physical reality which are coordinate-independent and is the rock on which present day emphasis on invariance principles will ultimately stand or fall.

[52] H. Everett, III, Rev. Mod. Phys. 29, 454 (1957).

[53] N. R. Graham, Ph.D. thesis, University of North Carolina (unpublished).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

requirements for a *random sequence with probabilities* $|c_n|^2$, to any arbitrary, but fixed, degree of accuracy, the resulting wave function is indistinguishable from the true final wave function in the limit $N \to \infty$. By "indistinguishable" we mean that the difference between it and the true wave function has vanishing norm.

The probability interpretation of quantum mechanics thus emerges from the formalism itself. Nonrandom memory sequences are "of measure zero" in the final superposition, in the limit $N \to \infty$. Each automaton (i.e., apparatus *cum* memory sequence) in the superposition sees the world obey the familiar quantum laws. However, there exists no outside agency which can designate which "branch" of the superposition is to be regarded as the *real* world. All are equally real, and yet each is unaware of the others. Thus if, within a given branch, an automaton, which has measured a given variable without changing it, subsequently checks his original observation, his memory sequence will not fail him. He will get his original value, and not that of some other branch. Moreover, if he communicates with another automaton who has simultaneously made the same measurement, their results will agree, which means that the two are in the same branch and that communication between different branches is impossible. The automaton therefore never feels himself split.

Everett's view of the world is a very natural one to adopt in the quantum theory of gravity, where one is accustomed to speak without embarassment of the "wave function of the universe." It is possible that Everett's view is not only natural but essential. For example, if the Hamiltonian constraint possesses only a single solution, so that the wave function for the universe is unique, then some conception like Everett's would appear to be needed in order to assess the physical significance of such uniqueness.[54]

In our discussion of the Friedmann model we assumed that the operator $\mathcal{H}+\mathcal{H}$ possesses a highly degenerate zero eigenvalue. How plausible is this assumption in the case of the actual world? In the case of the Friedmann model we were obliged to match the internal dynamics of the dust with that of the universe as a whole, with one hundred percent precision. Let us try to be a little more realistic. Suppose we replace the dust by a gas of noninteracting scalar bosons, but still maintain a rigid spherical geometry. Then we have an infinity of degrees of freedom. However, this infinity is *discrete*, because the universe is finite. Moreover, and this is important, there can never be more than a finite number of field quanta present in the state vector superposition, since the total energy (of the bosons) cannot be infinite. This is true even if the bosons are massless, since there is no infrared catastrophe in a finite world.

Now it is not at all difficult to verify that the Hamiltonian $\mathcal{H}$ in this case does not match $\mathcal{H}$ in any obviously commensurable way.[55] For each choice of boson quantum numbers $\mathcal{H}$ becomes a well-defined function of R, and the combination $\mathcal{H}+\mathcal{H}$ has a well-defined spectrum. But only by the sheerest accident does this spectrum include zero. All the evidence points to the fact that the complete spectrum of $\mathcal{H}+\mathcal{H}$, although discrete, is everywhere dense on the real line and does not condense into a set of finitely separated, infinitely degenerate levels. A similar situation holds with vector bosons and with fermions, and it seems hardly likely that the switching on of interactions between the particles will change the picture.

One might now suggest that we look for a way out of this predicament by relaxing the spherical rigidity restriction. However, this would merely correspond to the introduction of a gas of interacting tensor bosons, i.e., gravitons. It therefore appears that the same situation holds even in the general theory and that the Hamiltonian constraint of the real world may indeed have only one solution.[56]

If the state functional of the universe is unique how can we interpret it? In the case of the Friedmann model a single eigenfunction Ψ_n certainly has no resemblance to the real world, nor to any other reasonable world for that matter. A plot of $|\Psi_n|^2$ in the R-q plane looks like a lot of bumps separated from one another by a rectangular array of nodal lines, certainly nothing like a lissajous figure. However, suppose an extra term were added to the Hamiltonian $\mathcal{H}+\mathcal{H}$ which had the effect of strongly correlating the phase of the coherent particle clock with the phase of the universe, without changing either the (zero) eigenvalue or the quantum numbers. Then the Hamiltonian would no longer be separable and the nodal lines of $|\Psi_n|^2$ would no longer form a rectangular array. The bumps would instead tend to cluster around the Lissajous figure having the favored correlation, the figure itself now being somewhat distorted due to the correlation interaction, but still definitely recognizable.

The Hamiltonian of the real world is highly nonseparable, and there is a high degree of correlation among its infinity of modes. This must express itself as a kind of "condensation" of the state functional into components having many of the attributes of the quasiclassical Friedmann packets.[57] At the same time, because of the size of the universe, we know that the "Everett process" must be occurring on a lavish scale: The quasiclassical components of the universal state

[54] See J. A. Wheeler, The Monist 47, 40 (1962).

[55] M. Miketinac (private communication).

[56] The spectrum of $\mathcal{H}+\mathcal{H}$, or of $\mathcal{H}$ itself, can be shifted by the introduction of a "cosmological term" in the Einstein Lagrangian. If this spectrum is actually everywhere dense then we have the amusing result that a minute change in the cosmological constant can produce an enormous change in the zero-eigenvalue eigenvector and hence in the physical properties of the universe.

[57] If the state functional of the universe is unique then it is no longer possible or even meaningful to apply the inner-product definition (5.19) to the state functional as a whole. However, it might still be applied, in some reduced form, to its quasiclassical components.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

functional must be constantly splitting into a stupendous number of branches, all moving in parallel without interfering with one another except insofar as quantum Poincaré cycles allow rare anomalies to occur. According to the Everett interpretation each branch corresponds to a possible world-as-we-actually-see-it.

We have seen that the Friedmann packets in the R-q plane do not ultimately spread in "time"; every expansion-contraction cycle is exactly like every other. Unless some form of leakage to other channels occurs (e.g., transitions to different 3-space topologies) the same must be true for the real universe (assuming it to be closed and finite). In the absence of such channels there could be only *one* expansion-contraction cycle, repeated over and over again, like a movie film, throughout eternity, the monotony of which would be alleviated only by the infinite variety to be found among the multitude of simultaneous parallel worlds all executing the cycle together. Such a conclusion holds, in fact, regardless of whether the total state functional is unique or not.

A question naturally arises in regard to entropy. Within a given branch of the universal state functional the entropy would be observed (by appropriate automata) to increase with time.[55] It might be supposed that this increase would continue only during the expansion phase of the universe and that it would reverse itself during the contraction phase. This is not so, for one has only to remember that the length of a Poincaré cycle for even a small part of the universe is vastly longer than a rebound cycle, and hence except for a vanishingly small fraction of branches the entropy must continue to increase (at least locally) until final collapse is reached, at which point the very concepts of entropy and probability, as well as time itself, cease to have meaning.

However, if the operator $\mathfrak{IC}+\mathfrak{IC}$ is time-reversal invariant, and if its zero eigenvalue is nondegenerate, then the state functional of the universe is necessarily time-symmetric. This means that for every Everett branch in which entropy increases with time there must be another in which entropy decreases with time. To an observer in the second branch "time" in fact appears to be "flowing" in the opposite sense. Because of the extreme sensitivity of the state functional to slight changes in the operator $\mathfrak{IC}+\mathfrak{IC}$ (see Ref. 56) it is difficult to say how these conclusions must be modified if, as recent experiments suggest, the real world is not invariant under time reversal. However, the world being as complicated (and hence ergodic) as it is, it is still quite possible that there is no *preferred* direction in time. The ensemble of Everett branches in which time has a given direction of flow may very well be balanced by another ensemble in which time flows oppositely, so that reality as a whole possesses no over-all time orientation despite the absence of time-reversal invariance.

[55] Each branch corresponds to a pure state in the traditional sense. This does not, however, prevent the assignment of an effective entropy to it. For a sufficiently complicated system even a pure state may be assigned an entropy based on the coarse-grained properties of the state rather than on an ensemble average. In the classical theory this is illustrated by computer calculations of n-body systems. Even though the position and velocity of every body is known, the system as a whole possesses effective thermodynamical properties, the determination of which is in fact the goal of the computation.

ACKNOWLEDGMENTS

I wish to express my warmest gratitude for the kindness of Professor Robert Oppenheimer and Professor Carl Kaysen in extending to me the hospitality of the Institute for Advanced Study.

APPENDIX A: THE MANIFOLD M

M is defined as the 6-dimensional space of "points" (γ_{ij}) having as covariant and contravariant metric tensors, respectively, the expressions

$$G^{ijkl}=\tfrac{1}{2}\gamma^{1/2}(\gamma^{ik}\gamma^{jl}+\gamma^{il}\gamma^{jk}-2\gamma^{ij}\gamma^{kl}), \tag{A1}$$

$$G_{ijkl}=\tfrac{1}{2}\gamma^{-1/2}(\gamma_{ik}\gamma_{jl}+\gamma_{il}\gamma_{jk}-\gamma_{ij}\gamma_{kl}), \tag{A2}$$

satisfying

$$G_{ijab}G^{abkl}=\delta_{ij}{}^{kl}. \tag{A3}$$

Index pairs may, if desired, be mapped into single indices according to the rules

$$\gamma_{11}=\gamma^1, \quad \gamma_{22}=\gamma^2, \quad \gamma_{33}=\gamma^3,$$
$$\gamma_{23}=2^{-1/2}\gamma^4, \quad \gamma_{31}=2^{-1/2}\gamma^5, \quad \gamma_{12}=2^{-1/2}\gamma^6, \tag{A4}$$
$$\delta_{ij}{}^{kl}\to\delta^{\Gamma}{}_{\Delta}, \quad \Gamma,\Delta=1\cdots 6, \text{ etc.}$$

although this is seldom convenient or necessary.

By straightforward computation one may verify the variational law

$$G_{ijkl}\delta G^{ijkl}=-\gamma^{ij}\delta\gamma_{ij}, \tag{A5}$$

from which it may be inferred that

$$G\equiv\det(G^{ijkl})=-a\gamma^{-1}, \tag{A6}$$

where a is some constant. In the special case $\gamma_{ij}=\delta_{ij}$ one easily finds that the roots of G^{ijkl} are $-\frac{1}{2}$, 1, 1, 1, 1, 1, from which it follows that $a=\frac{1}{2}$ and that the signature of M is $-+++++$. The components γ_{ij} evidently are "good" coordinates in M as long as $\gamma\neq 0$.

If the ζ of Eq. (5.7) is chosen as a new coordinate then the surfaces of constant ζ have orthogonal trajectories whose tangent vectors are proportional to

$$\partial\zeta/\partial\gamma_{ij}=\tfrac{1}{4}\zeta\gamma^{ij}, \tag{A7}$$

or, in contravariant form,

$$G_{ijkl}\partial\zeta/\partial\gamma_{kl}=-\tfrac{1}{4}\zeta^{-1}\gamma_{ij}. \tag{A8}$$

If the orthogonal trajectories themselves are labeled by a set of five additional new coordinates ζ^A, $A=1\cdots 5$, then

$$\gamma^{ij}\partial\gamma_{ij}/\partial\zeta^A=4\zeta^{-1}\partial\zeta/\partial\zeta^A=0, \tag{A9}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

and, moreover, $\partial\gamma_{ij}/\partial\zeta$ must satisfy

$$G^{ijkl}\frac{\partial\gamma_{ij}}{\partial\zeta}\frac{\partial\gamma_{kl}}{\partial\zeta^A}=0. \tag{A10}$$

From these facts one may infer

$$G^{ijkl}\frac{\partial\gamma_{ij}}{\partial\zeta}\frac{\partial\gamma_{kl}}{\partial\zeta}=\left(G_{ijkl}\frac{\partial\zeta}{\partial\gamma_{ij}}\frac{\partial\zeta}{\partial\gamma_{kl}}\right)^{-1}=-1, \tag{A11}$$

$$\partial\gamma_{ij}/\partial\zeta=\tfrac{1}{3}\zeta^{-1}\gamma_{ij}, \tag{A12}$$

$$\gamma_{ij}\partial\zeta^A/\partial\gamma_{ij}=0, \tag{A13}$$

$$G^{ijkl}\frac{\partial\gamma_{ij}}{\partial\zeta^A}\frac{\partial\gamma_{kl}}{\partial\zeta^B}=\gamma^{1/2}\gamma^{ik}\gamma^{jl}\frac{\partial\gamma_{ij}}{\partial\zeta^A}\frac{\partial\gamma_{kl}}{\partial\zeta^B}, \tag{A14}$$

from which the metric (5.8) follows.

The above relations yield the following useful identities:

$$\mathrm{tr}(\gamma_{,A}\partial\zeta^B/\partial\gamma)=\delta_A{}^B, \tag{A15}$$

$$\mathrm{tr}(\gamma\partial\zeta^A/\partial\gamma)=0, \tag{A16}$$

$$\mathrm{tr}(\gamma^{-1}\gamma_{,A})=0, \tag{A17}$$

$$\mathrm{tr}[\gamma^{-1}(\gamma_{,AB}-\gamma_{,A}\gamma^{-1}\gamma_{,B})]=0, \tag{A18}$$

$$\frac{\partial\gamma_{ij}}{\partial\zeta^A}\frac{\partial\zeta^A}{\partial\gamma_{kl}}=\delta_{ij}{}^{kl}-\tfrac{1}{3}\gamma_{ij}\gamma^{kl}, \tag{A19}$$

$$\mathrm{tr}(\gamma_{,A}\mathbf{M})\,\mathrm{tr}(\mathbf{N}\partial\zeta^A/\partial\gamma)$$
$$=\tfrac{1}{2}\,\mathrm{tr}(\mathbf{MN}+\mathbf{MN}^{\sim})-\tfrac{1}{3}\,\mathrm{tr}(\gamma\mathbf{M})\,\mathrm{tr}(\gamma^{-1}\mathbf{N}), \tag{A20}$$

$$\mathrm{tr}(\gamma_{,AB}\mathbf{M})\,\mathrm{tr}(\mathbf{N}\partial\zeta^B/\partial\gamma)+\mathrm{tr}(\gamma_{,B}\mathbf{M})\,\mathrm{tr}[\mathbf{N}(\partial\zeta^B/\partial\gamma)_{,A}]$$
$$=-\tfrac{1}{3}\,\mathrm{tr}(\gamma_{,A}\mathbf{M})\,\mathrm{tr}(\gamma^{-1}\mathbf{N})$$
$$+\tfrac{1}{3}\,\mathrm{tr}(\gamma\mathbf{M})\,\mathrm{tr}(\gamma^{-1}\gamma_{,A}\gamma^{-1}\mathbf{N}), \tag{A21}$$

$$\mathrm{tr}[\gamma_{,A}\mathbf{M}(\partial\zeta^A/\partial\gamma)\mathbf{N}]$$
$$=\tfrac{1}{2}\,\mathrm{tr}(\mathbf{MN}^{\sim})+\tfrac{1}{2}\,\mathrm{tr}\mathbf{M}\,\mathrm{tr}\mathbf{N}-\tfrac{1}{3}\,\mathrm{tr}(\gamma\mathbf{M}\gamma^{-1}\mathbf{N}). \tag{A22}$$

Equations (A20) and (A22) follow from (A19), while Eqs. (A18) and (A21) are obtained from (A17) and (A20), respectively, by differentiation. $\mathbf{M}$ and $\mathbf{N}$ are arbitrary 3×3 matrices.

Using these identities and remembering the cyclic invariance of the trace, it is easy to compute the following:

$$\bar{G}^{AB}=\mathrm{tr}[\gamma(\partial\zeta^A/\partial\gamma)\gamma(\partial\zeta^A/\partial\gamma)], \tag{A23}$$
$$\bar{G}_{AC}\bar{G}^{CB}=\delta_A{}^B,$$

$$\Gamma_{ABC}\equiv\tfrac{1}{2}(\bar{G}_{AC,B}+\bar{G}_{BC,A}-\bar{G}_{AB,C})=\tfrac{1}{2}\mathrm{tr}[\gamma_{,C}\gamma^{-1}\times(-\gamma_{,A}\gamma^{-1}\gamma_{,B}-\gamma_{,B}\gamma^{-1}\gamma_{,A}+2\gamma_{,AB})\gamma^{-1}], \tag{A24}$$

$$\Gamma_{AB}{}^C\equiv\bar{G}^{CD}\Gamma_{ABD}=\mathrm{tr}[(-\gamma_{,A}\gamma^{-1}\gamma_{,B}+\gamma_{,AB})\partial\zeta^C/\partial\gamma], \tag{A25}$$

$$\bar{R}_{ABC}{}^D\equiv\Gamma_{BC}{}^D{}_{,A}-\Gamma_{AC}{}^D{}_{,B}+\Gamma_{BC}{}^E\Gamma_{AE}{}^A-\Gamma_{AC}{}^E\Gamma_{BE}{}^D=\mathrm{tr}[\gamma_{,C}\gamma^{-1}(\gamma_{,A}\gamma^{-1}\gamma_{,B}-\gamma_{,B}\gamma^{-1}\gamma_{,A})\times\partial\zeta^D/\partial\gamma], \tag{A26}$$

$$\bar{R}_{ABCD}\equiv\bar{R}_{ABC}{}^E\bar{G}_{ED}=\mathrm{tr}[\gamma^{-1}\gamma_{,D}\gamma^{-1}\gamma_{,C}\gamma^{-1}\times(\gamma_{,A}\gamma^{-1}\gamma_{,B}-\gamma_{,B}\gamma^{-1}\gamma_{,A})], \tag{A27}$$

$$\bar{R}_{AB}\equiv\bar{R}_{CAB}{}^C=-\tfrac{3}{2}\bar{G}_{AB}. \tag{A28}$$

The corresponding quantities in the full manifold M are

$$\Gamma_{AB}{}^C=\Gamma_{AB}{}^C, \tag{A29}$$

$$\Gamma_{AB}{}^0=(3/32)\zeta\bar{G}_{AB}, \tag{A30}$$

$$\Gamma_{A0}{}^B=\zeta^{-1}\delta_A{}^B, \tag{A31}$$

$$\Gamma_{A0}{}^0=\Gamma_{00}{}^A=\Gamma_{00}{}^0=0, \tag{A32}$$

$$R_{ABC}{}^D=\bar{R}_{ABC}{}^D-(3/32)(\bar{G}_{AC}\delta_B{}^D-\bar{G}_{BC}\delta_A{}^D),\quad+(\text{all other components vanish}) \tag{A33}$$

$$R_{AB}=\bar{R}_{AB}+\tfrac{3}{2}\bar{G}_{AB}=-(9/8)G_{AB}, \tag{A34}$$

$$R_{A0}=R_{00}=0, \tag{A35}$$

$${}^{(6)}R=-60\zeta^{-2}, \tag{A36}$$

where the index 0 is used for components in the direction of the "timelike" coordinate ζ.

The geodesic equation in $\bar{M}$ is obtained directly from (A25):

$$0=\frac{d^2\zeta^A}{ds^2}+\Gamma_{BC}{}^A\frac{d\zeta^B}{ds}\frac{d\zeta^C}{ds}=\mathrm{tr}\left[\frac{\partial\zeta^A}{\partial\gamma}\left(\frac{d^2\gamma}{ds^2}-\frac{d\gamma}{ds}\gamma^{-1}\frac{d\gamma}{ds}\right)\right]. \tag{A37}$$

This reduces to (5.11) upon multiplication with $\gamma_{,A}$ and use of (A17), (A18), and (A19).

When one stays within the manifold $\bar{M}$ it is convenient to map the matrices γ into matrices $\mathbf{a}$ of unit determinant:

$$\mathbf{a}\equiv\gamma^{-1/2}\gamma. \tag{A38}$$

In the solution (5.12) of the geodesic equation, γ may be replaced by $\mathbf{a}$ provided $\mathbf{M}$ is restricted to have unit determinant. To find the geodesic connecting two matrices $\mathbf{a}_1$ and $\mathbf{a}_2$, let $s=0$ at $\mathbf{a}_1$ and choose $\mathbf{M}$ in the form

$$\mathbf{M}=\mathbf{d}_1{}^{1/2}\mathbf{O}, \tag{A39}$$

where $\mathbf{O}$ is an orthogonal matrix which diagonalizes $\mathbf{a}_1$ and $\mathbf{d}_1{}^{1/2}$ is a diagonal square root of the resulting diagonal matrix. Then if s_{12} is the distance between $\mathbf{a}_1$ and $\mathbf{a}_2$ the matrix $\mathbf{N}$ satisfies

$$\mathbf{N}s_{12}=\ln(\mathbf{d}_1{}^{-1/2}\mathbf{O}\mathbf{a}_2\mathbf{O}^{\sim}\mathbf{d}_1{}^{-1/2}). \tag{A40}$$

From the condition $tr\mathbf{N}^2=1$ one obtains

$$s_{12}{}^2=\mathrm{tr}[\ln(\mathbf{d}_1{}^{-1/2}\mathbf{O}\mathbf{a}_2\mathbf{O}\mathbf{d}_1{}^{-1/2})]^2=\mathrm{tr}[\ln(\mathbf{a}_1{}^{-1}\mathbf{a}_2)]^2, \tag{A41}$$

which permits the matrix $\mathbf{N}$ itself to be determined. The logarithm of a matrix is an effectively unambiguous

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

concept, and the law of cyclic invariance of the trace applies to transcendental matrix functions as well as to rational functions. The uniqueness of the geodesic (5.12) is easily checked by noting that the matrix $\mathbf{O}$ is determined up to a transformation of the form $\mathbf{O}'=\mathbf{PO}$ where $\mathbf{P}$ is either a permutation matrix, if the roots of $\mathbf{a}_1$ are all distinct, or a more general orthogonal matrix, if some of the roots coincide. Such a transformation leaves (5.12) invariant. It is also easy to verify that it does not matter which of the eight possible square roots of $\mathbf{d}_1$ is chosen for $\mathbf{d}_1^{1/2}$.

The geodesic equations in M take the form

$$0=\frac{d^2\zeta}{ds^2}+(3/32)\zeta\left(\frac{d\delta}{ds}\right)^2, \tag{A42}$$

$$0=\frac{d^2\zeta^A}{ds^2}+\Gamma_{BC}{}^A\frac{d\zeta^B}{ds}\frac{d\zeta^C}{ds}+2\zeta^{-1}\frac{d\zeta^A}{ds}\frac{d\zeta}{ds}, \tag{A43}$$

where

$$\left(\frac{d\delta}{ds}\right)^2=\bar{G}_{AB}\frac{d\zeta^A}{ds}\frac{d\zeta^B}{ds}. \tag{A44}$$

Differentiating the latter equation and making use of (A43) multiplied by $\bar{G}_{AB}d\zeta^B/ds$, one finds

$$\frac{d^2\delta}{ds^2}=-\frac{2}{\zeta}\frac{d\delta}{ds}\frac{d\zeta}{ds}, \tag{A45}$$

which may be integrated to yield

$$\frac{d\delta}{ds}=\frac{\alpha}{\zeta^2}. \tag{A46}$$

α is an arbitrary integration constant which, without loss of generality, may be taken positive. When $\alpha\neq 0$ one may write

$$\frac{d\zeta^A}{d\delta}=\frac{\zeta^2}{\alpha}\frac{d\zeta^A}{ds}, \tag{A47}$$

which, in combination with (A43), yields (A37), showing that geodesics in M project onto geodesics in $\bar{M}$.

Substitution of (A46) into (A42) yields

$$\frac{d^2\zeta}{ds^2}+\frac{\kappa^2\alpha^2}{\zeta^3}=0,\quad \kappa\equiv(3/32)^{1/2}, \tag{A48}$$

which integrates to

$$d\zeta/ds=\pm(\kappa^2\alpha^2\zeta^{-2}-\beta)^{1/2}, \tag{A49}$$

where β is another integration constant. Using the metric (5.8) it is easy to verify that standard normalization for the affine parameter s is obtained by choosing $\beta=-1$, 0, or 1 according as the geodesic is timelike, null, or spacelike.

Timelike geodesics. In this case ζ must always increase (or decrease) with s. Therefore choosing the positive root in (A49) and the boundary conditions $\zeta(0)=0$, $\delta(\infty)=0$, one finds, upon setting $\beta=-1$ and integrating Eqs. (A46) and (A49),

$$\zeta(s)=[s(2\kappa\alpha+s)]^{1/2}=-\kappa\alpha\,\mathrm{csch}(\kappa\delta), \tag{A50}$$

$$\delta(s)=\frac{1}{2\kappa}\ln\frac{s}{2\kappa\alpha+s}. \tag{A51}$$

The ranges of the variables are

$$0\leqslant s<\infty,\quad -\infty\leqslant\delta<0,\quad 0\leqslant\zeta<\infty, \tag{A52}$$

and one sees that the geodesic strikes the frontier at $s=0$.

In terms of the matrix γ the above results may be expressed in the form

$$\gamma(s)=[-\kappa^2\alpha\,\mathrm{csch}(\kappa\delta)]^{4/3}\mathbf{M}\tilde{}\,e^{\mathbf{N}\ln\zeta}\mathbf{M}, \tag{A53}$$

where $\mathbf{N}$ is restricted as in (5.13) and $\mathbf{M}$ is now required to have unit determinant.

Null geodesics. In this case it is the constant α which serves to fix the scale of s. Setting $2\kappa\alpha=1$, $\beta=0$, and choosing the positive root in (A49), one finds, with the boundary condition $\zeta(0)=0$,

$$\zeta(s)=s^{1/2}. \tag{A54}$$

There is no preferred zero point for the arc length δ in $\mathbf{M}$. Hence the following integral of Eq. (A46) may be chosen:

$$\delta=(2\kappa)^{-1}\ln s. \tag{A55}$$

This yields

$$\zeta(s)=e^{\kappa\delta}, \tag{A56}$$

$$\gamma(s)=(\kappa e^{\kappa\delta})^{4/3}\mathbf{M}\tilde{}\,e^{\mathbf{N}\ln\zeta}\mathbf{M}. \tag{A57}$$

The ranges of the variables are

$$0\leqslant s<\infty,\quad -\infty\leqslant\delta<\infty,\quad 0\leqslant\zeta<\infty, \tag{A58}$$

and the geodesic is again seen to hit the frontier.

Spacelike geodesics. In this case there is a turning point at $\zeta=\kappa\alpha$ and both roots in (A49) can occur. Setting $\beta=1$, and choosing the boundary conditions $\zeta(0)=0$, $\delta(\kappa\alpha)=0$, with $s\geqslant 0$, one finds

$$\zeta(s)=[s(2\kappa\alpha-s)]^{1/2}=\kappa\alpha\,\mathrm{sech}(\kappa\delta), \tag{A59}$$

$$\delta(s)=\frac{1}{2\kappa}\ln\frac{s}{2\kappa\alpha-s}, \tag{A60}$$

$$\gamma(s)=[\kappa^2\alpha\,\mathrm{sech}(\kappa\delta)]^{4/3}\mathbf{M}\tilde{}\,e^{\mathbf{N}\ln\zeta}\mathbf{M}. \tag{A61}$$

The ranges of the variables are

$$0\leqslant s\leqslant 2\kappa\alpha,\quad -\infty\leqslant\delta\leqslant\infty,\quad 0\leqslant\zeta\leqslant\kappa\alpha, \tag{A62}$$

and the geodesic is seen to hit the frontier at both ends ($s=0$, 2κ).

Using the above results it is possible to obtain an expression for the "distance" to the frontier from any point in M. If γ is a fixed point and ξ is a contravariant vector at γ, then

$$\sigma(\gamma,\xi)=\frac{16}{3}\frac{[\mathrm{tr}\,\mathbf{X}^2-\frac{1}{2}(\mathrm{tr}\,\mathbf{X})^2]^{1/2}+(\frac{3}{2})^{1/2}\,\mathrm{tr}\,\mathbf{X}}{[\mathrm{tr}\,\mathbf{X}^2-\frac{1}{2}(\mathrm{tr}\,\mathbf{X})^2]^{1/2}-(\frac{3}{2})^{1/2}\,\mathrm{tr}\,\mathbf{X}}\gamma^{1/2}, \tag{A63}$$

$$\mathbf{X}=\gamma^{-1}\xi,$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where σ is one-half the square of the distance from γ to the frontier along the geodesic which starts from γ in the direction of ξ. The formula holds as long as $\mathrm{tr}\,\mathbf{X} < (\mathrm{tr}\,\mathbf{X}^2)^{1/2}$, so that the denominator is positive, for otherwise the geodesic escapes to infinity in the direction of ξ. σ takes on its minimum value, $-(16/3)\gamma^{1/2}$, when the geodesic is a path of pure dilation which arrives at the frontier at the point $\gamma=0$. The condition for this is $\mathrm{tr}\,\mathbf{X}^2=\frac{1}{3}(\mathrm{tr}\,\mathbf{X})^2$.

It should be noted that when the geodesic is not a path of pure dilation, it usually strikes the frontier at a point where some of the γ_{ij} become infinite, in spite of the fact that γ itself vanishes there. To see this, first observe that expressions (A53), (A57), and (A61) all have the limiting behavior

$$\gamma(s) \to (2\kappa^2\alpha e^{\kappa s})^{4/3}\mathbf{M}^{\sim}e^{\mathbf{N}s}\mathbf{M} \text{ as } s \to -\infty. \quad \text{(A64)}$$

Next note that in virtue of the conditions (5.13) $\mathbf{N}$ always has one root which is at least as negative as $-(\frac{1}{6})^{1/2}$. But $4\kappa/3=(\frac{1}{6})^{1/2}$. Therefore the only way in which a blow up of the exponential can be avoided in (A64) is for the roots of $\mathbf{N}$ to be precisely $-(\frac{1}{6})^{1/2}$, $-(\frac{1}{6})^{1/2}$, $(\frac{2}{3})^{1/2}$. Without loss of generality $\mathbf{N}$ may be chosen diagonal. The limiting form of $\gamma(s)$ in this special case is then

$$\gamma(0)=\mathbf{M}^{\sim}\begin{bmatrix}(2\kappa^2\alpha)^{4/3} & 0 & 0\\ 0 & (2\kappa^2\alpha)^{4/3} & 0\\ 0 & 0 & 0\end{bmatrix}\mathbf{M}. \quad \text{(A65)}$$

Since $\mathbf{M}$ and α are arbitrary (subject to $\det\mathbf{M}=1$) it follows that any singular symmetric matrix having an odd number of vanishing roots can be reached by a geodesic. Matrices having two vanishing roots can be reached (in a finite distance) from nonsingular points, but only along paths which suffer infinite absolute acceleration at the frontier.

It is not difficult to obtain an expression for the geodesic distance between two matrices γ_1 and γ_2 in M. The geometry of M may be summed up in the compact formula

$$ds^2=-d\zeta^2+\zeta^2 d\delta^2, \quad \text{(A66)}$$

which follows from (5.8). With the introduction of the variables

$$t=\zeta\cosh\kappa\delta, \quad x=\zeta\sinh\kappa\delta, \quad \text{(A67)}$$

this is converted to

$$ds^2=-dt^2+dx^2, \quad \text{(A68)}$$

which is formally just the line element of 2-dimensional Minkowski space. Hence

$$\begin{aligned}\sigma(\gamma_1,\gamma_2)\equiv\tfrac{1}{2}(s_{12})^2&=-\tfrac{1}{2}(t_1-t_2)^2+\tfrac{1}{2}(x_1-x_2)^2\\ &=\tfrac{1}{2}(2\zeta_1\zeta_2\cosh\kappa\delta_{12}-\zeta_1^2-\zeta_2^2)\\ &=(16/3)\{2(\gamma_1\gamma_2)^{1/4}\cosh[(3/32)^{1/2}\delta_{12}]\\ &\qquad -\gamma_1^{1/2}-\gamma_2^{1/2}\}, \quad \text{(A69)}\end{aligned}$$

where δ_{12} is given by (A41).

APPENDIX B: THE MANIFOLDS $M^{\infty 1}$ AND $\mathfrak{M}$

In discussing these manifolds it will be convenient to use an abbreviated notation which avoids the necessity of writing integral signs or excessive numbers of indices bearing various numbers of primes. This notation will be applicable to completely general manifolds and, in fact, will be used again in the following paper of this series in quite a different context, thus providing additional justification for its introduction here.

The functions $\gamma_{ij}(\mathbf{x})$ will be replaced by the symbol φ^i. More precisely, the symbol γ is replaced by φ, and the quintuple (i,j,x^1,x^2,x^3) by the single index i. In general applications the φ's may constitute either a finite discrete set of real numbers or, as here, a set of functions or "fields." When the index i has a continuous character, the summation convention for repeated indices will be understood to include integrations over the continuous labels for which it stands. In this Appendix no restriction will be placed on the range of values which the indices can assume.

The φ's are "coordinates" in a manifold ($M^{\infty 1}$ in the present case) on which a group acts. Group elements will be denoted by barred letters $\bar{x}$, $\bar{y}$, etc. and their components in some coordinate system in the group space will be denoted by $\bar{x}^\alpha$, $\bar{y}^\beta$, etc. For example, the 3-dimensional general coordinate transformation group may be coordinatized by the functions $\bar{x}^i(\mathbf{x})$ which define the coordinate transformation $x^i \to \bar{x}^i$. In the condensed notation the quadruplet (i,x^1,x^2,x^3) gets replaced by the single index α.

The multiplication table of the group defines a set of function(al)s $F^\alpha[\bar{y},\bar{x}]$ satisfying

$$F^\alpha[\bar{y},\bar{x}]=(\bar{y}\bar{x})^\alpha. \quad \text{(B1)}$$

For example, in the case of the coordinate transformation group this functional has the form

$$F^{i,x}[\bar{y},\bar{x}]=\bar{y}^i(\bar{\mathbf{x}}(\mathbf{x})).$$

By virtue of the group postulates $F^\alpha[\bar{x},\bar{y}]$ also satisfies the following fundamental identities:

$$F^\alpha[\bar{x},e]=F^\alpha[e,\bar{x}]=\bar{x}^\alpha, \quad \text{(B2)}$$

$$F^\alpha[\bar{x},\bar{x}^{-1}]=F^\alpha[\bar{x}^{-1},\bar{x}]=e^\alpha, \quad \text{(B3)}$$

$$F^\alpha[\bar{z},\bar{y}\bar{x}]=F^\alpha[\bar{z}\bar{y},\bar{x}], \quad \text{(B4)}$$

where e is the identity element of the group and $\bar{x}^{-1}$ denotes the inverse of $\bar{x}$. In the case of the coordinate transformation group, we have $e^i(\mathbf{x})=x^i$.

Instead of dealing directly with $F^\alpha[\bar{y},\bar{x}]$, one more often makes use of a set of auxiliary funcation(al)s, together with the structure constants of the group. These are defined, respectively, by

$$L^\alpha{}_\beta[\bar{x}]\equiv(\delta F^\alpha[\bar{y},\bar{x}]/\delta\bar{y}^\beta)_{\bar{y}=e}, \quad \text{(B5)}$$

$$c^\alpha{}_{\beta\gamma}\equiv(\delta^2F^\alpha[\bar{y},\bar{x}]/\delta\bar{y}^\gamma\delta\bar{x}^\beta)_{\bar{x}=\bar{y}=e}. \quad \text{(B6)}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

In the case of the coordinate transformation group one finds $L^i{}_{j'}[\bar{x}]=\delta^i{}_j\delta(\bar{\mathbf{x}}(\mathbf{x}),\mathbf{x}')$, while the structure constants are as given in Eq. (4.17). In the case of finite-dimensional Lie groups the functional derivatives in (B5) and (B6) become ordinary derivatives.

By repeatedly differentiating Eqs. (B2), (B3), (B4) and setting various elements equal to the identity, a number of important relations can be established. Among them we cite the following:

$$L^\alpha{}_\beta[e]=\delta^\alpha{}_\beta, \tag{B7}$$

$$L^{-1\alpha}{}_{\beta,\epsilon}-L^{-1\alpha}{}_{\epsilon,\beta}=-c^\alpha{}_{\beta\gamma}L^{-1\beta}{}_\delta L^{-1\gamma}{}_\epsilon, \tag{B8}$$

$$c^\alpha{}_{\beta\delta}c^\gamma{}_{\epsilon\phi}+c^\alpha{}_{\gamma\epsilon}c^\gamma{}_{\delta\beta}+c^\alpha{}_{\delta\epsilon}c^\gamma{}_{\beta\gamma}=0. \tag{B9}$$

In Eq. (B8) the arguments of the funcation(al)s have been suppressed, and differentiation is denoted by a comma. $L^{-1\alpha}{}_\beta$ denotes the matrix inverse to $L^\alpha{}_\beta$. In the case of the coordinate transformation group it is given by $L^{-1i}{}_{j'}[\bar{x}]=\delta^i{}_j\delta(\mathbf{x},\bar{\mathbf{x}}(\mathbf{x}'))\partial(\bar{\mathbf{x}}')/\partial(\mathbf{x}')$.

As a result of the action of the group the variables φ^i suffer a transformation which may be expressed in the form

$$\varphi'^i=\Phi^i[\bar{x},\varphi], \tag{B10}$$

where the function(al)s $\Phi^i[\bar{x},\varphi]$ satisfy the identities

$$\Phi^i[e,\varphi]=\varphi^i, \tag{B11}$$

$$\Phi^i[\bar{y}\bar{x},\varphi]=\Phi^i[\bar{y},\Phi[\bar{x},\varphi\,]]. \tag{B12}$$

Differentiation of (B12) leads to

$$\Phi^i{}_{,\alpha}[\bar{x},\varphi]=R^i{}_\beta[\Phi[\bar{x},\varphi]]L^{-1\beta}{}_\alpha[\bar{x}], \tag{B13}$$

where

$$R^i{}_\alpha[\varphi]\equiv\Phi^i{}_{,\alpha}[e,\varphi]. \tag{B14}$$

The function(al)s $R^i{}_\alpha$ appear in the law of transformation of the φ's under infinitesimal group operations. Under the action of a group element having the coordinates $\epsilon^\alpha+\delta\xi^\alpha$, where the $\delta\xi$'s are infinitesimal, the φ's suffer the change

$$\delta\varphi^i=R^i{}_\alpha\delta\xi^\alpha. \tag{B15}$$

(Functional arguments are again suppressed.) For example, under the infinitesimal coordinate transformation $\bar{x}^i=x^i+\delta\xi^i$, the 3-metric γ_{ij} suffers the change

$$\delta\gamma_{ij}=\int R_{ij\mathbf{x}'}\delta\xi^{\mathbf{x}'}d^3x', \tag{B16}$$

where

$$R_{ij\mathbf{x}'}\equiv-\gamma_{ij,k}\delta(\mathbf{x},\mathbf{x}')-\gamma_{kj}\delta_{,i}(\mathbf{x},\mathbf{x}')-\gamma_{ik}\delta_{,j}(\mathbf{x},\mathbf{x}') \tag{B17a}$$

$$=-\delta_{i\mathbf{x}',j}-\delta_{j\mathbf{x}',i}, \tag{B17b}$$

$$\delta_{ij'}\equiv\gamma_{ik}\delta^k{}_{j'}. \tag{B18}$$

If the transformation laws (B10) and (B14) are linear as in this case, then $R^i{}_{\alpha,jk}=0$. (The reader should avoid confusing differentiation with respect to the x's in the explicit notation and functional differentiation with respect to the φ's in the compact notation.)

With these preliminaries out of the way the question of imposing a metric on the manifold of φ's may now be considered. Let such a metric be denoted by $g_{ij}[\varphi]$. If the group is to generate isometric motions in the manifold then this metric must satisfy Killing's equation:

$$g_{ij,k}\delta\varphi^k+g_{kj}\delta\varphi^k{}_{,i}+g_{ik}\delta\varphi^k{}_{,j}=0, \tag{B19}$$

with $\delta\varphi^i$ given by Eq. (B15). It is not difficult to see that this equation may be regarded as a group transformation law for g_{ij}:

$$\delta g_{ij}\equiv g_{ij,k}\delta\varphi^k=-g_{kj}R^k{}_{\alpha,i}\delta\xi^\alpha-g_{ik}R^k{}_{\alpha,j}\delta\xi^\alpha. \tag{B20}$$

When the transformation law (B15) is linear and homogeneous, then

$$\delta\varphi^i=R^i{}_{\alpha,j}\varphi^j\delta\xi^\alpha, \tag{B21}$$

and Eq. (B20) says simply that g_{ij} must transform contragrediently to the Kronecker product $\varphi^i\varphi^j$. This is a necessary and sufficient condition for the isometry of group operations.

Now let $d\varphi^i$ be an arbitrary displacement, with the corresponding "arc length" ds given by

$$ds^2=g_{ij}d\varphi^id\varphi^j. \tag{B22}$$

If $d\varphi^i$ happens to be orthogonal to the *orbit* of φ under the group then it satisfies the condition

$$R^j{}_\alpha g_{ij}d\varphi^j=0, \tag{B23}$$

and (B22) gives directly the distance between neighboring orbits. In the case of the manifold $M^{\infty 3}$, with the metric (6.5) and the transformation law (B16), (B17), the condition (B23) takes the form

$$\gamma^{jk}(d\gamma_{ij,k}-d\gamma_{jk,i})=0. \tag{B24}$$

More generally, the distance between orbφ and orb$(\varphi+d\varphi)$ is given by

$$ds^2=g_{ij}\bar{d}\varphi^i\bar{d}\varphi^j, \tag{B25}$$

where $\bar{d}\varphi^i$ is the projection of $d\varphi^i$ normal to the orbit:

$$\bar{d}\varphi^i=(\delta^i{}_j-R^i{}_\alpha\gamma^{\alpha\beta}R^k{}_\beta g_{kj})d\varphi^j, \tag{B26}$$

$$\gamma_{\alpha\gamma}\gamma^{\gamma\beta}=\delta_\alpha{}^\beta, \tag{B27}$$

$$\gamma_{\alpha\beta}\equiv g_{ij}R^i{}_\alpha R^j{}_\beta. \tag{B28}$$

When the indices α, β include continuous labels the "matrix" $\gamma_{\alpha\beta}$ is tyipcally a differential operator (sum of differentiated δ functions) and its inverse $\gamma^{\alpha\beta}$ is a Green's function.

Equation (B25) may be written in the alternative forms

$$ds^2=\bar{g}_{ij}d\varphi^id\varphi^j=\bar{g}_{ij}\bar{d}\varphi^i\bar{d}\varphi^j, \tag{B29}$$

where

$$\bar{g}_{ij}\equiv g_{ij}-g_{ik}R^k{}_\alpha\gamma^{\alpha\beta}R^l{}_\beta g_{lj}. \tag{B30}$$

$\bar{g}_{ij}$ is the metric in the manifold of orbits, and the question arises how *it* transforms under the group. This question is answered by establishing the following

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

transformation laws:

$$\delta R^i{}_\alpha \equiv R^i{}_{\alpha,j}\delta\varphi^j = (R^i{}_{\beta,j}R^i{}_\alpha - c^\gamma{}_{\beta\alpha}R^i{}_\gamma)\delta\xi^\beta, \quad \text{(B31)}$$

$$\delta\gamma_{\alpha\beta} \equiv \gamma_{\alpha\beta,i}\delta\varphi^i = -(c^\gamma{}_{\delta\alpha}\gamma_{\gamma\beta} + c^\gamma{}_{\delta\beta}\gamma_{\alpha\gamma})\delta\xi^\delta, \quad \text{(B32)}$$

$$\delta\gamma^{\alpha\beta} \equiv \gamma^{\alpha\beta}{}_{,i}\delta\varphi^i = (c^\alpha{}_{\delta\gamma}\gamma^{\gamma\beta} + c^\beta{}_{\delta\gamma}\gamma^{\alpha\gamma})\delta\xi^\delta. \quad \text{(B33)}$$

Equations (B32) and (B33) are corollaries of (B19), (B27), and (B31), while (B31) itself is a consequence of the identity

$$R^i{}_{\alpha,j}R^j{}_\beta - R^i{}_{\beta,j}R^j{}_\alpha = R^i{}_\gamma c^\gamma{}_{\alpha\beta}, \quad \text{(B34)}$$

which is obtained by differentiating Eq. (B13) with respect to $\bar{x}^\beta$, setting $\bar{x} = e$, antisymmetrizing in α and β, and making use of (B8). With the aid of (B31) and (B33) it is straightforward to show that $\bar{g}_{ij}$ transforms just like g_{ij}. This means that group operations are isometries of $\bar{g}_{ij}$ just as they are of g_{ij}, and suggests that $\bar{g}_{ij}$ is effectively a function of orbφ alone. In order to make this fact explicit a coordinatization of the orbit manifold will be introduced.

This may be accomplished by first introducing a hypersurface in the φ manifold, defined by a set of simultaneous equations

$$f_\alpha[\varphi] = 0, \quad \text{(B35)}$$

where the index α ranges over the same continuum (or discrete set, as the case may be) as the group indices. The only requirement on the hypersurface is that it intersect the orbit of every point contained in (at least) some finite portion of the φ manifold. A coordinate system is then laid down in this hypersurface, with the coordinates denoted by z^A. If the hypersurface has been carefully chosen each orbit will intersect it in a single point, and the z's at that point may be used to label the orbit itself. For example, in the manifold $M^{\infty 4}$ one may choose for the equations (B35) the harmonic condition $(\gamma^{1/2}\gamma^{ij})_{,j} = 0$; then any three of the functions $\varphi^{AB}(\eta)$ of Eq. (5.3) may be chosen as the z's.

A general point in the φ manifold will be reached by moving off the hypersurface along (i.e., within) an orbit thus:

$$\varphi^i[\bar{x}, z] = \Phi^i[\bar{x}, \varphi_0[z]], \quad \text{(B36)}$$

where $\varphi_0{}^i[z]$ is the starting point on the hypersurface. The group coordinates $\bar{x}^\alpha$ together with the z's provide a new labeling scheme for the points of the φ manifold, and the task before us is to compute the metrics g_{ij} and $\bar{g}_{ij}$ in this new coordinate system. For this purpose we shall need the relations

$$\varphi^i{}_{,\alpha} = R^i{}_\beta[\varphi]L^{-1\beta}{}_\alpha[\bar{x}], \quad \text{(B37)}$$

$$\varphi^i{}_{,\alpha A} = R^i{}_{\beta,j}[\varphi]\varphi^j{}_{,A}L^{-1\beta}{}_\alpha[\bar{x}], \quad \text{(B38)}$$

$$\varphi^i{}_{,\alpha\beta} = R^i{}_{\gamma,j}[\varphi]R^j{}_\delta[\varphi]L^{-1\gamma}{}_\alpha[\bar{x}]L^{-1\delta}{}_\beta[\bar{x}] + R^i{}_\gamma[\varphi]L^{-1\gamma}{}_{\alpha,\beta}[\bar{x}], \quad \text{(B39)}$$

which are obtained by applying Eq. (B13) to (B36). In the work which follows the arguments $\bar{x}$, φ, and z will be suppressed.

It is straightforward to compute

$$g_{\alpha\beta} \equiv g_{ij}\varphi^i{}_{,\alpha}\varphi^j{}_{,\beta} = \gamma_{\gamma\delta}L^{-1\gamma}{}_\alpha L^{-1\delta}{}_\beta, \quad \text{(B40)}$$

$$g_{\alpha A} = g_{A\alpha} \equiv g_{ij}\varphi^i{}_{,\alpha}\varphi^j{}_{,A} = g_{ij}R^i{}_\beta L^{-1\beta}{}_\alpha \varphi^j{}_{,A}, \quad \text{(B41)}$$

$$g_{AB} \equiv g_{ij}\varphi^i{}_{,A}\varphi^j{}_{,B} = \bar{g}_{AB} + g_{A\alpha}\bar{g}^{\alpha\beta}g_{\beta B}, \quad \text{(B42)}$$

where

$$\bar{g}_{AB} \equiv \bar{g}_{ij}\varphi^i{}_{,A}\varphi^j{}_{,B}, \quad \text{(B43)}$$

$$\bar{g}^{\alpha\beta} \equiv L^\alpha{}_\gamma L^\beta{}_\delta \gamma^{\gamma\delta}, \quad g_{\alpha\gamma}\bar{g}^{\gamma\beta} = \delta_\alpha{}^\beta. \quad \text{(B44)}$$

One then readily verifies that the contravariant metric, with components $g^{\alpha\beta}$, $g^{\alpha A}(= g^{A\alpha})$, g^{AB}, is given by

$$g^{\alpha\beta} = \bar{g}^{\alpha\beta} + \bar{g}^{\alpha\gamma}g_{\gamma A}g^{AB}g_{B\delta}\bar{g}^{\delta\beta}, \quad \text{(B45)}$$

$$g^{\alpha A} = -\bar{g}^{\alpha\beta}g_{\beta B}g^{BA}, \quad \text{(B46)}$$

$$\bar{g}_{AC}g^{CB} = \delta_A{}^B. \quad \text{(B47)}$$

It is now easy to show that g_{AB} and $\bar{g}_{AB}$ (and hence g^{AB}) are independent of the x's. Thus, using (B37) and (B38) one finds

$$g_{AB,\alpha} = g_{ij,k}\varphi^i{}_{,A}\varphi^j{}_{,B}\varphi^k{}_{,\alpha} + g_{ij}(\varphi^i{}_{,A\alpha}\varphi^j{}_{,B} + \varphi^i{}_{,A}\varphi^j{}_{,B\alpha}) = (g_{ij,k}R^k{}_\beta + g_{kj}R^k{}_{\beta,i} + g_{ik}R^k{}_{\beta,j}) \times \varphi^i{}_{,A}\varphi^j{}_{,B}L^{-1\beta}{}_\alpha. \quad \text{(B48)}$$

But this expression vanishes by virtue of the group transformation law for g_{ij} [Eq. (B20)]. Since $\bar{g}_{ij}$ obeys the same transformation law it follows that

$$\bar{g}_{AB,\alpha} = 0. \quad \text{(B49)}$$

That is, the metric $\bar{g}_{AB}$ of the orbit manifold depends only on the z's, as was expected.

For the study of geodesics in the φ manifold the following derivatives will also be needed:

$$g_{\alpha\beta,\gamma} = g_{ij,k}\varphi^i{}_{,\alpha}\varphi^j{}_{,\beta}\varphi^k{}_{,\gamma} + g_{ij}(\varphi^i{}_{,\alpha\gamma}\varphi^j{}_{,\beta} + \varphi^i{}_{,\alpha}\varphi^j{}_{,\beta\gamma}) = g_{\alpha\delta}L^\delta{}_\epsilon L^{-1\epsilon}{}_{\gamma,\beta} + g_{\beta\delta}L^\delta{}_\epsilon L^{-1\epsilon}{}_{\gamma,\alpha}, \quad \text{(B50)}$$

$$g_{\alpha A,\beta} = g_{ij,k}\varphi^i{}_{,\alpha}\varphi^j{}_{,A}\varphi^k{}_{,\beta} + g_{ij}(\varphi^i{}_{,\alpha\beta}\varphi^j{}_{,A} + \varphi^i{}_{,\alpha}\varphi^j{}_{,A\beta}) = g_{A\gamma}L^\gamma{}_\delta L^{-1\delta}{}_{\beta,\alpha}. \quad \text{(B51)}$$

These are obtained with the aid of (B19), (B37), (B38), and (B39).

The geodesic equations in the φ manifold may be written in the form

$$0 = g_{\alpha\beta}\frac{d^2\bar{x}^\beta}{ds^2} + g_{\alpha A}\frac{d^2z^A}{ds^2} + \Gamma_{\beta\gamma\alpha}\frac{d\bar{x}^\beta}{ds}\frac{d\bar{x}^\gamma}{ds} + 2\Gamma_{\beta A\alpha}\frac{d\bar{x}^\beta}{ds}\frac{dz^A}{ds} + \Gamma_{AB\alpha}\frac{dz^A}{ds}\frac{dz^B}{ds}, \quad \text{(B52)}$$

$$0 = g_{A\alpha}\frac{d^2\bar{x}^\alpha}{ds^2} + g_{AB}\frac{d^2z^B}{ds^2} + \Gamma_{\alpha\beta A}\frac{d\bar{x}^\alpha}{ds}\frac{d\bar{x}^\beta}{ds} + 2\Gamma_{\alpha BA}\frac{d\bar{x}^\alpha}{ds}\frac{dz^B}{ds} + \Gamma_{BCA}\frac{dz^B}{ds}\frac{dz^C}{ds}, \quad \text{(B53)}$$

where the Γ's are the Christoffel symbols. Multiplying

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

(B52) by $g_{A\dot{\alpha}}\bar{g}^{\dot{\alpha}a}$, subtracting the result from (B53) and using the fact that $g_{AB,a}=0$, one obtains

$$0=\bar{g}_{AB}\frac{d^2z^B}{ds^2}+\Gamma_{BCA}\frac{dz^B}{ds}\frac{dz^C}{ds}+\tfrac{1}{2}[g_{\alpha A,\beta}+g_{\beta A,\alpha}-g_{\alpha\beta,A}-g_{A\dot{\delta}}\bar{g}^{\dot{\delta}\gamma}(g_{\alpha\gamma,\beta}+g_{\beta\gamma,\alpha}-g_{\alpha\beta,\gamma})]\frac{d\bar{x}^\alpha}{ds}\frac{d\bar{x}^\beta}{ds}$$

$$+[g_{\alpha A,B}-g_{\alpha B,A}-g_{A\gamma}\bar{g}^{\gamma\beta}(g_{\alpha\beta,B}+g_{\beta B,\alpha}-g_{\alpha B,\beta})]\frac{d\bar{x}^\alpha}{ds}\frac{dz^B}{ds}+\tfrac{1}{2}[(g_{B\alpha}\bar{g}^{\alpha\beta}g_{\beta A})_{,C}+(g_{C\alpha}\bar{g}^{\alpha\beta}g_{\beta A})_{,B}-(g_{B\alpha}\bar{g}^{\alpha\beta}g_{\beta C})_{,A}$$

$$-g_{A\dot{\beta}}\bar{g}^{\dot{\beta}\alpha}(g_{B\alpha,C}+g_{C\alpha,B})]\frac{dz^B}{ds}\frac{dz^C}{ds}, \quad \text{(B54)}$$

where the Γ's are the Christoffel symbols of the orbit manifold. The terms of this equation can be regrouped by judicious use of identities such as $g_{\alpha\beta,A}=(g_{\alpha\gamma}\bar{g}^{\gamma\delta}g_{\delta\beta})_{,A}$ and $\bar{g}^{\gamma\delta}{}_{,A}g_{\delta\beta}=-\bar{g}^{\gamma\delta}g_{\delta\beta,A}$ and by replacing derivatives of the form $g_{\alpha\beta,\gamma}$ and $g_{\alpha A,\beta}$ by their expressions (B50), (B51). The final useful result is

$$0=\bar{g}_{AB}\frac{d^2z^B}{ds^2}+\Gamma_{BCA}\frac{dz^B}{ds}\frac{dz^C}{ds}+[(g_{A\gamma}\bar{g}^{\gamma\alpha})_{,C}-(g_{C\gamma}\bar{g}^{\gamma\alpha})_{,A}]\frac{dz^C}{ds}\left(g_{\alpha\beta}\frac{d\bar{x}^\beta}{ds}+g_{\alpha B}\frac{dz^B}{ds}\right)$$

$$+\tfrac{1}{2}\bar{g}^{\gamma\delta}{}_{,A}\left(g_{\gamma\alpha}\frac{d\bar{x}^\alpha}{ds}+g_{\gamma B}\frac{dz^B}{ds}\right)\left(g_{\delta\beta}\frac{d\bar{x}^\beta}{ds}+g_{\delta C}\frac{dz^C}{ds}\right)+g_{A\gamma}\bar{g}^{\gamma\beta}L^{\dot{\delta}}{}_{\epsilon}(L^{-1\epsilon}{}_{\beta,\alpha}-L^{-1\epsilon}{}_{\alpha,\beta})\frac{d\bar{x}^\alpha}{ds}\left(g_{\dot{\delta}\dot{\zeta}}\frac{d\bar{x}^\zeta}{ds}+g_{\dot{\delta}B}\frac{dz^B}{ds}\right). \quad \text{(B55)}$$

Now suppose the geodesic intersects one of the orbits orthogonally. The condition for this is [cf. Eq. (B23)]

$$R^i{}_{jk}g_{ir}d\varphi^j/ds=0, \quad \text{(B56)}$$

which, when multiplied by $L^{-1\beta}{}_\alpha$, yields

$$g_{\alpha\beta}\frac{d\bar{x}^\beta}{ds}+g_{\alpha A}\frac{dz^A}{ds}=0. \quad \text{(B57)}$$

When this condition is satisfied we have

$$g_{AB}\frac{dz^A}{ds}\frac{dz^B}{ds}+2g_{\alpha A}\frac{d\bar{x}^\alpha}{ds}\frac{dz^A}{ds}+g_{\alpha\beta}\frac{d\bar{x}^\alpha}{ds}\frac{d\bar{x}^\beta}{ds}=\bar{g}_{AB}\frac{dz^A}{ds}\frac{dz^B}{ds}, \quad \text{(B58)}$$

and hence

$$ds^2=d\bar{s}^2, \quad \text{(B59)}$$

so that the arc length in the φ manifold becomes the same as in the orbit manifold. Moreover, by virtue of (B55) it follows that the z's in this case satisfy also the geodesic equation in the orbit manifold,

$$\bar{g}_{AB}\frac{d^2z^B}{d\bar{s}^2}+\Gamma_{BCA}\frac{dz^B}{d\bar{s}}\frac{dz^C}{d\bar{s}}=0, \quad \text{(B60)}$$

provided the orthogonality condition (B57) is maintained along the entire length of the geodesic. But this is an immediate consequence of Eqs. (B50), (B51), and (B52), for by differentiating the left-hand side of (B57) with respect to s, one obtains

$$g_{\alpha\beta}\frac{d^2\bar{x}^\beta}{ds^2}+g_{\alpha A}\frac{d^2z^A}{ds^2}+g_{\alpha\beta,\gamma}\frac{d\bar{x}^\beta}{ds}\frac{d\bar{x}^\gamma}{ds}+(g_{\alpha\beta,A}+g_{\alpha A,\beta})\frac{d\bar{x}^\beta}{ds}\frac{dz^A}{ds}+g_{\alpha A,B}\frac{dz^A}{ds}\frac{dz^B}{ds}$$

$$=\tfrac{1}{2}g_{\beta\gamma,\alpha}\frac{d\bar{x}^\beta}{ds}\frac{d\bar{x}^\gamma}{ds}+g_{\beta A,\alpha}\frac{d\bar{x}^\beta}{ds}\frac{dz^A}{ds}=\left(g_{\dot{\delta}\beta}\frac{d\bar{x}^\beta}{ds}+g_{\dot{\delta}A}\frac{dz^A}{ds}\right)L^{\dot{\delta}}{}_{\epsilon}L^{-1\epsilon}{}_{\alpha,\gamma}\frac{d\bar{x}^\gamma}{ds}, \quad \text{(B61)}$$

which vanishes by virtue of (B57) itself. Therefore, if the geodesic intersects one orbit orthogonally then it intersects every orbit in its path orthogonally, and, moreover, it traces out a geodesic curve in the orbit manifold.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ERRATA

Quantum Theory of Gravity

I. The Canonical Theory.

p. 1119, Column 2, line 11, read "(3.8b)" for "(3.8a)".

p. 1121, Column 2, line 21 should read

" = $\lim(2\pi)^{-1}(1+\varepsilon)[f_\varepsilon(x-\sqrt{\varepsilon}) + f_\varepsilon(x+\sqrt{\varepsilon}) - 2\varepsilon f_\varepsilon(x)/(1+\varepsilon)]$".

p. 1122, Column 2. Add the word "the" at the end of the eleventh line from the bottom.

p. 1124, Column 1, line 7, read "extent" for "estent".

p. 1125, Column 2, line 18, omit the word "to".

p. 1125, Column 2, add the word "unit" at the end of the seventeenth line from the bottom.

p. 1126, Column 2, line 8, read "to the frontier" for "in the frontier".

p. 1131, Column 1, line 16, read "ultimately" for "utlimately".

p. 1136, Column 1, sixth line from bottom, read "cycloidal" for "cyclodial".

p. 1146, Column 1, line 13, read "function (al)s" for "funcation (al)s".

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

A NEW TYPE OF ISOTROPIC COSMOLOGICAL MODELS WITHOUT SINGULARITY

A.A. STAROBINSKY

Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, England [1] *and The Landau Institute for Theoretical Physics, The Academy of Sciences, Moscow, 117334, USSR* [2]

Received 11 January 1980

The Einstein equations with quantum one-loop contributions of conformally covariant matter fields are shown to admit a class of nonsingular isotropic homogeneous solutions that correspond to a picture of the Universe being initially in the most symmetric (de Sitter) state.

It is well known that many solutions of the classical Einstein equations, in particular, the Friedmann–Robertson–Walker isotropic homogeneous cosmological model, contain singularities and cannot be analytically continued beyond them. So a fundamental cosmological problem arises: what was there before the "big bang", i.e. before the stage of classical expansion of the Universe?

One might think that quantum-gravitational effects will somehow change the situation. At present there exists a particular case where they can be calculated in closed form; it is the effects due to the interaction of quantum free matter fields with a classical gravitational field. The one-loop approximation is exact in this case. It looks reasonable to solve the Einstein equations with quantum corrections in this approximation and see if there exists any physically interesting nonsingular solution.

Let us consider the simplest case: the self-consistent evolution of an isotropic homogeneous classical gravitational background interacting with a number of massless conformally covariant quantum fields with different spin values. The space–time metric has the form:

$$ds^2 = dt^2 - a^2(t)\, d\Sigma_K^2(r, \theta, \varphi)\,, \qquad (1)$$

$K = +1, 0, -1$ corresponds to a closed, flat and open Friedmann–Robertson–Walker model, respectively. $a(t)$ should be determined from the Einstein equations with quantum contributions in the right-hand-side [‡1]:

$$R_{ik} - \tfrac{1}{2} g_{ik} R = 8\pi G \langle T_{ik} \rangle\,. \qquad (2)$$

There is no massless particle creation and nonlocal vacuum polarization in the conformally flat metric (1), so $\langle T_{ik} \rangle$ consists only of local terms which arise in the process of regularization (see, e.g., refs. [3,4]):

$$\begin{aligned}\langle T_{ik} \rangle = {} & (k_2/2880\pi^2)(R_i{}^l R_{kl} - \tfrac{2}{3} R R_{ik} \\ & - \tfrac{1}{2} g_{ik} R_{lm} R^{lm} + \tfrac{1}{4} g_{ik} R^2) \\ & + (k_3/2880\pi^2)\tfrac{1}{6}(2R_{;i;k} - 2g_{ik} R^{;l}{}_{;l} \\ & - 2R R_{ik} + \tfrac{1}{2} g_{ik} R^2)\,. \end{aligned} \qquad (3)$$

The constants k_2 and k_3 are sums of contributions from quantum fields with different spin values.

Solutions of eqs. (2), (3) have physical sense only as long as $|R_{iklm} R^{iklm}| \ll t_g^{-4}$, $t_g = \sqrt{G}$, because if this is not the case higher-loop corrections due to quantized gravitational and matter interactions cannot be neglected. So if we succeed in finding at least one solution which is nonsingular at all t and does not violate

[1] The work in Cambridge University was supported by The Nuffield Foundation.

[2] Permanent address (also address after 25 December, 1979).

[‡1] $\hbar = c = 1$ is assumed throughout this paper. The Landau–Lifshitz sign conventions are used. In particular, compared with ref. [1], R_{iklm}, R_{ik} and R (but not T) have the opposite sign, and with ref. [2], R and T have the opposite sign.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the above mentioned condition then we shall obtain one possible answer to the fundamental question stated in the first paragraph of this paper. It is worth noting that the evolution of the Universe need not follow a "generic" solution, it may well be described just by this unique one, at least initially.

Eqs. (2), (3) were first considered in ref. [1] and then investigated in detail in ref. [2] in the case $K = 0$. The conclusion was that they have no nonsingular solutions. We shall show that the only reason for this result is that an additional term describing classical isotropic radiation (i.e. a certain number of initially present free massless particles with an isotropic distribution function) was added to $\langle T_{ik} \rangle$ in ref. [2]. If this term is absent, so that there are no particles in the initial state of the Universe, then there exists a one-parameter family of nonsingular solutions of eqs. (2), (3). These solutions are not of the "bouncing" type [‡2], they describe the Universe beginning its evolution from the de Sitter state with constant curvature. After that, the Friedmann expansion with the effective equation of state $p = 0$ takes place. In terms of the full quantum theory of gravity, the initial most symmetric state can probably be thought of as an eigenstate of the curvature. This scenario of the beginning of the Universe is the extreme opposite of Misner's initial "chaos".

The 00 component of eq. (2) with $\langle T_{ik} \rangle$ given by eq. (3) has the form:

$$\frac{\dot{a}^2 + K}{a^2} = \frac{1}{H^2}\left(\frac{\dot{a}^2 + K}{a^2}\right)^2 - \frac{1}{M^2}\left[2\frac{\dot{a}\dddot{a}}{a^2} - \frac{\ddot{a}^2}{a^2} + 2\frac{\ddot{a}\dot{a}^2}{a^3} - 3\left(\frac{\dot{a}}{a}\right)^4 - 2K\frac{\dot{a}^2}{a^4} + \frac{K^2}{a^4}\right], \tag{4}$$

where $H^2 = 360\pi/Gk_2$, $M^2 = -360\pi/Gk_3$. We assume that $k_2 > 0$ and (in contrast to ref. [2]) $k_3 < 0$. The former condition is necessary for the existence of the de Sitter solution, the latter follows from the requirement of stability of the classical solutions of the Einstein equations. Both conditions are met, e.g., in case of the quantum electromagnetic field. Further one should admit that $Ht_g \ll 1$, $Mt_g \ll 1$ to justify the one-loop approximation. This can be achieved if the number of quantum fields is large enough.

Let $u = a^2\dot{a}^2$, $v = u + Ka^2$. Then, instead of eq. (4), we have:

$$v = \frac{v^2}{H^2a^4} - \frac{1}{M^2}\left(\frac{vv''}{a^2} - \frac{2vv'}{a^3} - \frac{v'^2}{4a^2} - Kv'' + \frac{3Kv'}{a}\right), \tag{5}$$

where a prime denotes a derivative with respect to a.

Another useful representation of eq. (4) can be obtained by letting $f = u^{3/4}$, $\xi = (12)^{-3/4}a^3$. Then

$$d^2f/d\xi^2 + M^2\xi^{-2/3}f^{-1/3} - M^2f/12H^2\xi^2 + K[\sqrt{12}M^2f^{-5/3} - 3^{-1/2}\xi^{-4/3}f^{-1/3}(M^2/H^2 + 1)] + K^2\xi^{-2/3}f^{-5/3}(1 - M^2/H^2) = 0. \tag{6}$$

Eqs. (4)–(6) have the partial de Sitter solution:

$$v(a) = H^2a^4, \quad f(\xi) = 12^{3/8}\xi^{1/2}(\sqrt{12}H^2\xi^{2/3} - K)^{3/4},$$

$$a(t) = H^{-1}\cosh Ht, \quad K = +1,$$

$$= a_0\exp(Ht), \quad K = 0,$$

$$= H^{-1}\sinh Ht, \quad K = -1. \tag{7}$$

In the case $K = -1$, the space–time is regular at the horizon $t = 0$ and can be continued beyond it.

The existence of such a self-consistent solution was first mentioned in ref. [6]. The nonsingular solution (7) in itself has nothing to do with our low-curvature Universe but it proves to be unstable both to the past (that means that nonsingular solutions are not generic) and, that is more important, to the future [‡3]. Let

$$v(a) = H^2a^4[1 + \delta(a)], \quad |\delta| \ll 1. \tag{8}$$

To a first approximation in δ, eq. (5) becomes

$$\delta''\left(1 - \frac{K}{H^2a^2}\right) + \frac{\delta'}{a}\left(4 - \frac{5K}{H^2a^2}\right) - \frac{M^2\delta}{H^2a^2} = 0. \tag{9}$$

A one-parameter family of solutions remaining close to (7) in the past has the form:

$$\delta = \delta_0 a^{\alpha - 3/2}, \quad \alpha = (\tfrac{9}{4} + M^2/H^2)^{1/2}, \quad K = 0,$$

$$\delta = \delta_0 F(\tfrac{3}{4} + \tfrac{1}{2}\alpha, \tfrac{3}{4} - \tfrac{1}{2}\alpha, 3; -a^2H^2), \quad K = -1,$$

$$\delta = \delta_0(aH)^{-(\alpha+3/2)}F(\tfrac{3}{4} + \tfrac{1}{2}\alpha, -\tfrac{5}{4} + \tfrac{1}{2}\alpha, 1 + \alpha; 1/a^2H^2), \quad K = +1, \tag{10}$$

where F is the hypergeometric function (in the case $K = +1$, the solution is displayed for the stage $-\infty <$

[‡2] "Bouncing" nonsingular solutions can be obtained in the model with a one-loop contribution of a nonconformal massless quantum field [5].

[‡3] The solution (7) is future stable in the case $k_3 > 0$.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

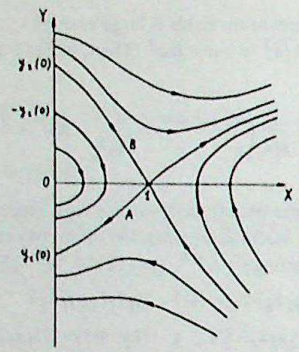


Fig. 1. The phase plane of eq. (12). Arrows show the direction of increasing ξ and a.

$t < 0$ only). None of the solutions (10) remain close to (7) for $t \to \infty$.

If we want our solution to match the parameters of the real Universe then stage (7) should be long enough: $Ht_0 \gg 1$, where t_0 is the moment of transition to a Friedmann stage ($\delta(t_0) \approx 1$). This enables us to neglect spatial curvature terms (K terms) in eq. (6) when investigating the transition region.

So let $K = 0$. By making the substitution

$$f = (12H^2)^{3/4}\xi x\,, \quad y = \xi\, \mathrm{d}x/\mathrm{d}\xi \tag{11}$$

in eq. (6), we obtain the following equation:

$$\mathrm{d}y/\mathrm{d}x = M^2(x - x^{-1/3})/12H^2 y - 1\,. \tag{12}$$

Its phase diagram is shown in fig. 1. The unstable saddle point (1,0) is the de Sitter solution (7). All curves going to infinity correspond to singular solutions. The family of nonsingular solutions of eq. (6) correspond to a solution of eq. (12) that moves from the point (1,0) along the curve A. Let $y_1(x)$ be the equation of curve A and $y_2(x)$ be the one of curve B.

The $x = 0$ axis corresponds to points of inflexion of $a(t)$ ($\dot{a} = \ddot{a} = 0$, $\dddot{a} \neq 0$). A phase curve reaching the y-axis at some point $(0, y_0)$ continues its way from the point $(0, -y_0)$. So, to establish that the extension of the phase curve A will always remain in the part of the phase plane confined by the curves A and B and the y-axis (so the corresponding solution will be nonsingular) one should only prove that $y_2(0) > |y_1(0)|$. This is easily achieved by multiplying eq. (12) by y and subtracting the resulting equation for $y_1(x)$ from the one for $y_2(x)$, which gives

$$y_2^2(0) - y_1^2(0) = 2\int_0^1 \mathrm{d}x\, [|y_1(x)| + y_2(x)] > 0\,. \tag{13}$$

The behaviour of the solution at the Friedmann stage can be obtained from the analysis of oscillations which was carried out in ref. [5] in a more general case. Adapting the results to the case involved, we find that for $t \gg \max(M^{-1}, H^{-1})$:

$$a(t) = a_1 t^{2/3}[1 + (2/3Mt)\sin M(t - t_1)]\,. \tag{14}$$

Note that the corresponding oscillations of $f(\xi)$ are not linear and cannot be considered as small perturbations.

One can say that a number of heavy scalar quasiparticles with mass M and zero momentum are created which determine the law of the Friedmann expansion. In the model considered, these "scalarons" are stable. But if one generalizes the model and includes one-loop contributions of massive quantum fields then scalarons will decay into pairs of massive particles and have a finite life time $\tau = 6M/G\mu^4$ in the case of neutral scalar particles with mass μ and $\tau = 6/GM\mu^2$ in the case of spinor particles (the conditions $\mu \ll M, t_g^{-1}$ are assumed). It is interesting that the decay of scalarons into pairs of gravitons appears to be strongly suppressed if Nariai's corrections [7] to the Lifshitz equation for gravitational waves in an isotropic background are taken into account.

The important property of all nonsingular models with the initial superdense de Sitter state is that, as shown in ref. [8], such a large amount of relic gravitational waves is generated by one-loop processes in these models (in particular, in the range $1-10^{-5}$ Hz) that predictions of the semiclassical theory and the very existence of this state can be experimentally verified in the near future. Adopting such a model, one should also call for some mechanism of baryon-number generation because initial symmetry requires zero initial values of all charges.

The author acknowledges the hospitality of Professor S.W. Hawking and his whole group in DAMTP, Cambridge University, where this paper was written.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

References

[1] P.C.W. Davies, Phys. Lett. 68B (1977) 402.
[2] M.V. Fischetti, J.B. Hartle and B.L. Hu, Phys. Rev. D20 (1979) 1757.
[3] P.C.W. Davies, S.A. Fulling, S.M. Christensen and T.S. Bunch, Ann. Phys. 109 (1977) 108.
[4] T.S. Bunch and P.C.W. Davies, Proc. R. Soc. London A356 (1977) 569.
[5] V.Ts. Gurovich and A.A. Starobinsky, Zh. Eksp. Teor. Fiz. 77 (1979) 1699.
[6] J.S. Dowker and R. Critchley, Phys. Rev. D13 (1976) 3224.
[7] H. Nariai, Prog. Theor. Phys. 49 (1973) 165.
[8] A.A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. 30 (1979), to be published.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

COSMOLOGICAL BARYON-NUMBER DOMAIN STRUCTURE AND THE FIRST ORDER PHASE TRANSITION OF A VACUUM

Katsuhiko SATO
Nordita, DK-2100 Copenhagen Ø, Denmark
and Department of Physics, Kyoto University, Kyoto, Japan [1]

Received 4 February 1980
Revised manuscript received 24 July 1980

If *CP*-nonconservation arises from spontaneous symmetry breaking in the very early universe, the universe will have a domain structure of baryon number. We propose a model of the early universe in which domains are stretched exponentially and the radius of the domains is much greater than that of the horizon of the standard big bang model, provided that the grand unified theory undergoes a first order phase transition. If the size of the stretched domains is sufficiently big to avoid pair annihilations of baryon and antibaryon domains, the difficulties of the baryon symmetric universe may be removed.

It has been suggested very recently that the baryon number asymmetry of the universe is a natural consequence of the baryon- and *CP*-nonconserving interactions in the modern grand unified theory, GUT [1]. This theory is very interesting from the point of view that the gauge theory with spontaneous symmetry breaking developed recently also accounts for one of the most important problems of cosmology. In this theory [1], we must break the symmetry of the physical law in a special direction in order to explain this asymmetry of the universe, i.e., the sign of the *CP*-violating interactions would be the same everywhere in space. However, if we believe that basic laws of physics should always be symmetrical, then *CP*-conservation should be broken spontaneously [2–4]. This assumption, however, predicts a baryon-number symmetric universe with local fluctuations of the baryon number. Brown and Stecker [3] considered optimistically that baryon number domains grow to astrophysical size, such as super clusters of galaxies, by the merging process proposed by Omnes [5]. But the size of *CP*-domains is, at most, the particle horizon when the *CP*-conservation is broken spontaneously [3]. It seems to be very difficult that these small domains grow sufficiently [6].

The purpose of the present paper is to propose a model of the early universe in which the size of the *CP*-domains is stretched very much, provided that the phase transition in the GUT is of first order.

First we assume that the *CP*-domain structure of the universe was created by the spontaneous breaking of the *CP*-conservation at the cosmic time t_{cp} and the radius of the domains is of the order of the particle horizon at this time as discussed by Brown and Stecker [3], namely the radius is $2t_{cp}$ provided that the radiation energy density is dominant. In the present work, the *CP*-breaking time t_{cp} is a free parameter. However, it is reasonable to presume that the time t_{cp} is equal to or earlier than the time of the phase transition in the GUT, but is later than the Planck time.

Secondly, we assume that the GUT phase transition is of first order and the transition time scale of the phase transition is longer than the characteristic time scale of the cosmic expansion at this time. Kirzhnits and Linde [7] and Linde [8] investigated the vacuum phase transition in the abelian Higgs model. In this model, the vacuum phase transition is of first order if the coupling constant of the four-body

[1] Present address.

0 031–9163/81/0000–0000/$ 02.50 © North-Holland Publishing Company

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

self-interaction λ is less than the square of the gauge coupling constant g, i.e., $\lambda < g^2$. They showed, moreover, that the transition time scale becomes longer and longer with the decrease of the coupling constant λ due to the radiative correction. In the limit $\lambda \to 3g^4/32\pi^2$, the transition time scale becomes infinite. The vacuum phase transition in the GUT is more complicated than that of the abelian Higgs model and has not yet been studied in detail. But, recently, Guth and Tye [9] exhibited a GUT phase transition model in which the phase transition is of first order. It is, therefore, possible that the GUT phase transition is of first order [7–9].

The effective potential of the vacuum energy density $V(\phi, T)$ which gives the first order phase transition has a local minimum at $\phi = 0$ even after the temperature becomes less than the critical temperature T_c (see fig. 1), where ϕ is an order parameter of the vacuum or a Higgs field and T is the cosmic temperature. When the temperature becomes less than the critical temperature T_c, the nucleation of the bubbles of the stable vacuum ($\phi \neq 0$) begins. The bubbles expand with almost light velocity [10] and the phase transition finishes when all the space is covered by bubbles. In the model of the abelian Higgs model with $g^2 > \lambda > 3g^4/16\pi^2$, however, the local minimum at $\phi = 0$ disappears when the temperature becomes less than the temperature T_{c1} (see fig. 1). Linde [8]

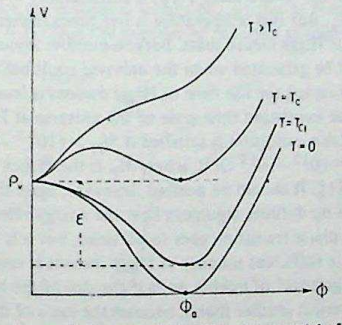


Fig. 1. Schematic diagram of the effective potential of the vacuum energy density. When the cosmic temperature T becomes less than T_c, the phase transition begins. When the temperature becomes equal to T_{c1}, the local minimum at $\phi = 0$ disappears and the phase transition finishes instantaneously at this time.

considered that the phase transition finishes instantaneously at this temperature because the nucleation rate becomes infinite. In this case, the size of bubbles when they coalesce and terminate the phase transition will be much smaller than the particle horizon. In the model with $\lambda < 3g^4/16\pi^2$, however, the local minimum at $\phi = 0$ never disappears even if the temperature becomes zero. In the latter case, domains whose size is comparable to the particle horizon will be created if the nucleation rate is small. (Expansion and coalescence of bubbles in the expanding universe have been discussed in detail in ref. [11]).

In the present work, we take the phase transition model similar to that of the abelian Higgs model with $g^2 > \lambda > 3g^4/16\pi^2$ in order to avoid the complicated discussion about the thermalization of the kinetic energy of the bubble walls (see the later discussion). We simply assume that the phase transition finishes instantaneously when the cosmic temperature becomes equal to T_{c1}, which is an arbitrary parameter in the present work. (Nucleations of bubbles for $T > T_{c1}$ are neglected.)

Now let us derive the expansion law of the universe before the finishing time of the phase transition. As discussed by Bludman and Ruderman [12] and Linde [13], the vacuum before the phase transition must have a huge energy density if the vacuum energy density at present is zero. Let the energy density of the vacuum before the phase transition be ρ_v (see fig. 1).

The total energy density of the universe before the phase transition is the sum of ρ_v and the radiation energy density $\rho_r \equiv S\pi^2 T^4/15$, where the statistical weight in units of photons S is of order 100. If we assume the universe is a homogeneous Minkowski space, the change of the cosmic scale factor R is described as

$$\dot{R}^2/R^2 = l^{-2} + bl^2/R^4, \tag{1}$$

where l is the characteristic scale factor of this model,

$$l \equiv (8\pi G\rho_v/3)^{-1/2} \tag{2}$$

and b is a constant which is defined as $b \equiv \rho_r R^4/\rho_v l^4$. The solution of this equation is

$$\begin{aligned} R(t) &= b^{1/4} l \sinh^{1/2}(2t/l) \\ &= b^{1/4} l (2t/l)^{1/2}, \qquad \text{for } 2t \ll l, \\ &= (b^{1/4} l/\sqrt{2}) \exp(t/l), \qquad \text{for } 2t \gg l. \end{aligned} \tag{3}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

After the cosmic time $l/2$, the vacuum energy density becomes dominant when the phase transition has not yet finished. The finishing time t_1 of the phase transition is obtained from eq. (3) as a function of the temperature T_{c1},

$$t_1 = \tfrac{1}{2} l \,\mathrm{arcsinh}\,(T_v^4/ST_{c1}^4)^{1/2}, \tag{4}$$

where we have defined T_v as

$$\tfrac{1}{15}\pi^2 T_v^4 \equiv \rho_v. \tag{5}$$

If $T_{c1} < S^{-1/4}\, T_v$, the finishing time t_1 is later than the time $l/2$ and the universe expands exponentially in the period between $l/2$ and t_1. Eq. (4) reduces to

$$t_1 = l \ln[2^{1/2} S^{-1/4} (T_v/T_{c1})],$$

provided $T_{c1} \ll S^{-1/4}\, T_v$.

The radius of the *CP*-domains is also stretched exponentially by the cosmic expansion. The radius at the finishing time $L(t_1)$ is calculated from eq. (3) as

$$L(t_1) = L(t_{cp}) \sinh(2t_1/l)/\sinh(2t_{cp}/l), \tag{6}$$

where $L(t_{cp})$ is the initial radius of the *CP*-domains. As is discussed already, the size $L(t_{cp})$ is assumed to be the size of the particle horizon $2t_{cp}$, if the radiation energy density is dominant at the time of *CP*-breaking, i.e., $t_{cp} < l/2$. Provided that $t_{cp} < l/2 < t_1$, eq. (6) reduces to

$$L(t_1) = (t_{cp} l)^{1/2} \exp(t_1/l)$$

$$= S^{-1/4}(2t_{cp} l)^{1/2}(T_v/T_{c1}). \tag{7}$$

When the phase transition finishes, the energy density difference between two vacua (the stable and the meta stable vacua) ϵ (see fig. 1) is released and the universe is strongly heated up if the energy is thermalized. In the case of the first order phase transition, however, the released energy ϵ goes to the kinetic energy of the bubble walls initially [10]. This kinetic energy will be thermalized when the bubbles collide and coalesce with each other. (But it has not yet been investigated in detail how the energy is thermalized.) However, if the radius of the bubbles is much greater than l, a time longer than the timescale of the cosmic expansion is necessary for the radiation created at the bubble wall to go to the center of bubbles.

We have taken the phase transition model in which the local minimum disappears at the temperature T_{c1}. In this model, most of the radii of bubbles are much smaller than l.

If the energy released by the phase transition is thermalized instantaneously, the temperature immediately after the phase transition $T(t_1)$ is given as

$$T(t_1) = (15\epsilon/\pi^2 S)^{1/4} \approx S^{-1/4}\, T_v. \tag{8}$$

In the second equality, we assumed that ϵ is almost equal to ρ_v, because the effective potential $V(\phi, T_{c1})$ is almost the same as $V(\phi, 0)$ provided $T_{c1}/T_c \ll 1$. If the temperature $T(t_1)$ is comparable to or higher than the mass of the lepto-quark boson M_X or the Higgs meson mass M_H, the baryon-number asymmetry is created in each *CP*-domain by the decays of these particles in the course of the following expansion of the universe. Using the abelian Higgs model, we now estimate whether the cosmic temperature $T(t_1)$ is sufficiently high to produce the lepto-quark bosons or the Higgs mesons. In this model, $M_X = g\mu/\sqrt{\lambda}$ and $M_H = \sqrt{2}\mu$. In the limit $\lambda \to 3g^4/16\pi^2$, T_{c1} and ρ_v reduce to zero and $\mu^4/2\lambda$, respectively [7,8]. Then we obtain the relation between the temperature and these masses as

$$T(t_1) = 0.1(100/S)^{1/4} M_X$$

$$= 5(100/S)^{1/4}(0.1/g) M_H. \tag{9}$$

The creation of baryon-number asymmetry by the decays of lepto-quark bosons seems to be very difficult, because the temperature $T(t_1)$ is considerably lower than M_X. But this temperature is five times higher than the Higgs meson mass. Baryon-number asymmetry will be generated when the universe cools below M_H, provided the life time of Higgs mesons is longer than the expansion time scale of the universe at $T = M_H$ (this condition is satisfied if $M_H > (10^{-7}-10^{-6}) \times M_P = 10^{12}-10^{13}$ GeV, where M_P is the Planck mass) [1]. It should be noticed, however, again that there is no definite argument how the energy released by the phase transition goes to particles, but it is likely that a sufficient number of Higgs mesons is created as "fragments" of bubble walls if the size of the bubbles is much smaller than l, because the ratio of the Compton wave length of the Higgs meson M_H^{-1} to l is of the order of $10^{-3}(M_X/10^{15}\ \mathrm{GeV})(0.1/g)$.

The subsequent evolution of the universe after the phase transition in the present model is essentially the same as that of the standard big-bang model (SBB

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

model) [14] if pair annihilations of the baryon and antibaryon domains are negligible. In the SBB model without the vacuum energy density, the size of the particle horizon when the cosmic temperature $T = S^{-1/4} T_\nu$, which is the temperature immediately after the heating-up in the present model [eq. (8)], is l, because the particle horizon in the SBB model is given as $(8\pi^3 GST^4/45)^{-1/2}$ [14]. The ratio of the size of the domains immediately after the phase transition to the size of the particle horizon of the SBB model at the same temperature $S^{-1/4} T_\nu$ is obtained from eq. (7) as

$$r = (t_{cp}/l)^{1/2} \exp(t_1/l) = S^{-1/4} (2t_{cp}/l)^{1/2} (T_\nu/T_{c1}). \quad (10)$$

If $t_{cp} = l/2$ and $t_1 \gg l$, the domain size is much bigger than the particle horizon of the SBB model. Of course this does not conflict with causality, because the particle horizon in the present model has been stretched even more than the domain size (see fig. 2).

Now we discuss the case that the domains are extremely stretched and that the radius of the domains at present is comparable to the mean separation distance between super clusters of galaxies, d. The radius of the domain at present L_0 is estimated by the ratio of the present cosmic radiation temperature T_0 to $T(t_1)$,

$$L_0 = L(t_1)(S/S_0)^{1/3} T(t_1)/T_0, \quad (11)$$

where the statistical weight of particles at present S_0 is $1 + 21/22$ if the number of neutrino species is three,

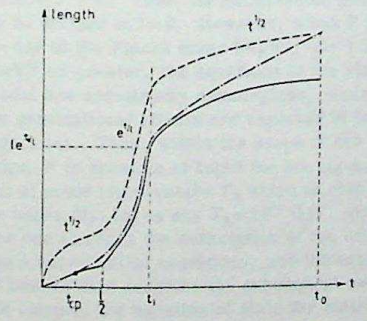


Fig. 2. Schematic diagram of the changes of the cosmic scale factor $R(t)$, ---; the radius of a domain $L(t)$, ——; and the particle horizon of this cosmic model, -.-.-.

i.e., ν_e, ν_μ and ν_τ [14]. From the condition that the present baryon number domains should be greater than d, i.e. $L_0 > d$, we obtain the following inequality substituting eqs. (7), (8) and (11),

$$T_{c1} < 3 \times 10^2 (2/S_0)^{1/3} (100/S)^{1/6} \times (3\,\mathrm{K}/T_0)(100\,\mathrm{Mpc}/d)(2t_{cp}/l)^{1/2} \quad \mathrm{eV}. \quad (12)$$

As Kirzhnits and Linde [7] and Linde [8] have shown, the temperature T_{c1} reduces to zero in the limiting case $\lambda \to 3g^4/16\pi^2$ in the abelian Higgs model. In principle we cannot rule out the possibility of the case $T_{c1} < 3 \times 10^2$ eV in the GUT phase transition, but this is obviously very unnatural. At present, there are no definite arguments to estimate T_{c1} in the GUT phase transition, but it may be reasonable to presume that the lower limit T_{c1} is of the order of the masses of the weak bosons $M_{W^\pm,Z} \approx 10^2$ GeV, because this is the smallest quantity with the dimension of energy in the GUT (neglecting fermion masses). If $T_{c1} = 10^2$ GeV, $T_\nu = 10^{15}$ GeV ($l = 3 \times 10^{-36}$ s) and $S = 100$, the ratio of the size of the domains immediately after the phase transition to the particle horizon of the SBB model at the same temperature r is $3 \times 10^{12} \times (2t_{cp}/l)^{1/2}$ from eq. (10). Even if the CP-breaking arises at the Planck time $G^{-1/2} = 5 \times 10^{-44}$ s instead of the time of the GUT phase transition $l/2$, the ratio r is of order 10^9. Unfortunately we cannot say clearly that the size of the domains such as $r = 10^9$ is sufficiently big in order for the domains to grow to the size of super clusters of galaxies by the merging process [5], because the critical size of the baryon domains (which determines whether the domains can grow or are annihilated) is very ambiguous [3,5,6].

In conclusion, we showed that CP-domains are stretched exponentially and very big baryon-number domains are created compared with those of the standard model if the GUT phase transition is of first order. In principle, the domains can grow to the size of the mean separation distance between super clusters of galaxies, d, but the very unnatural assumption $T_{c1} < 3 \times 10^2$ eV is necessary. Under the reasonable assumption $T_{c1} > M_{Z,W^\pm} \approx 10^2$ GeV, the merging process of domains [5] must work in order for the domains to grow to the size d even if the size of the domains is extremely stretched. We may, however, conclude that the present model makes the baryon-number symmetric universe with the baryon-number

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

domain structure more likely than the original model proposed by Brown and Stecker [3]. But if the merging process does not work as stressed by Steigman [6], the hypothesis of spontaneous breaking of *CP*-conservation would be ruled out even if the GUT phase transition is of first order.

The author thanks M.B. Einhorn, R. Epstein, H.B. Nielsen and C.J. Pethick for stimulating discussions and kind and critical comments.

References

[1] M. Yoshimura, Phys. Rev. Lett. 41 (1978) 281;
A.Yu. Ignatiev, N.V. Krasnikov, V.A. Kuzim and A.N. Tavkhelidze, Phys. Lett. 76B (1978) 436;
S. Dimopoulos and L. Susskind, Phys. Rev. D18 (1978) 4500;
D. Toussaint, S.B. Treiman, F. Wilczek and A. Zee, Phys. Rev. D19 (1979) 1036;
S. Dimopoulos and L. Susskind, Phys. Lett. 81B (1979) 416;
J. Ellis, M.K. Gaillard and D.V. Nanopoulos. Phys. Lett. 80B (1979) 360;
S. Weinberg, Phys. Rev. Lett. 42 (1979) 850;
J.D. Barrow, Mon. Not. R. Astron. Soc. 189 (1979) 23L.
[2] Y.B. Zeldovich, L.B. Okun and I.Yu. Kobzarev, Sov. Phys. JETP 40 (1974) 1.
[3] R.W. Brown and F.W. Stecker, Phys. Rev. Lett. 43 (1979) 315.
[4] J.J. Aly, Cambridge preprint (1979).
[5] R. Omnes, Phys. Rep. 3C (1972) 1.
[6] G. Steigman, Ann. Rev. Astrophys. 14 (1976) 339.
[7] D.A. Kirzhnits and A.D. Linde, Ann. Phys. 101 (1976) 195.
[8] A.D. Linde, Rep. Prog. Phys. 42 (1979) 389.
[9] A.H. Guth and S.-H.H. Tye, Phys. Rev. Lett. 44 (1980) 631.
[10] S. Coleman, Phys. Rev. D15 (1977) 2929.
[11] K. Sato, preprint (1980).
[12] S.A. Bludman and M.A. Ruderman, Phys. Rev. Lett. 38 (1977) 255.
[13] A.D. Linde, JETP Lett. 19 (1974) 183.
[14] S. Weinberg, Gravitation and cosmology (Whiley, New York, 1972).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PHYSICAL REVIEW D VOLUME 23, NUMBER 2 15 JANUARY 1981

Inflationary universe: A possible solution to the horizon and flatness problems

Alan H. Guth*
Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305
(Received 11 August 1980)

The standard model of hot big-bang cosmology requires initial conditions which are problematic in two ways: (1) The early universe is assumed to be highly homogeneous, in spite of the fact that separated regions were causally disconnected (horizon problem); and (2) the initial value of the Hubble constant must be fine tuned to extraordinary accuracy to produce a universe as flat (i.e., near critical mass density) as the one we see today (flatness problem). These problems would disappear if, in its early history, the universe supercooled to temperatures 28 or more orders of magnitude below the critical temperature for some phase transition. A huge expansion factor would then result from a period of exponential growth, and the entropy of the universe would be multiplied by a huge factor when the latent heat is released. Such a scenario is completely natural in the context of grand unified models of elementary-particle interactions. In such models, the supercooling is also relevant to the problem of monopole suppression. Unfortunately, the scenario seems to lead to some unacceptable consequences, so modifications must be sought.

I. INTRODUCTION: THE HORIZON AND FLATNESS PROBLEMS

The standard model of hot big-bang cosmology relies on the assumption of initial conditions which are very puzzling in two ways which I will explain below. The purpose of this paper is to suggest a modified scenario which avoids both of these puzzles.

By "standard model," I refer to an adiabatically expanding radiation-dominated universe described by a Robertson-Walker metric. Details will be given in Sec. II.

Before explaining the puzzles, I would first like to clarify my notion of "initial conditions." The standard model has a singularity which is conventionally taken to be at time $t=0$. As $t\to 0$, the temperature $T\to\infty$. Thus, no initial-value problem can be defined at $t=0$. However, when T is of the order of the Planck mass ($M_P \equiv 1/\sqrt{G} = 1.22\times 10^{19}$ GeV)[1] or greater, the equations of the standard model are undoubtedly meaningless, since quantum gravitational effects are expected to become essential. Thus, within the scope of our knowledge, it is sensible to begin the hot big-bang scenario at some temperature T_0 which is comfortably below M_P; let us say $T_0 = 10^{17}$ GeV. At this time one can take the description of the universe as a set of initial conditions, and the equations of motion then describe the subsequent evolution. Of course, the equation of state for matter at these temperatures is not really known, but one can make various hypotheses and pursue the consequences.

In the standard model, the initial universe is taken to be homogeneous and isotropic, and filled with a gas of effectively massless particles in thermal equilibrium at temperature T_0. The initial value of the Hubble expansion "constant" H is taken to be H_0, and the model universe is then completely described.

Now I can explain the puzzles. The first is the well-known horizon problem.[2-4] The initial universe is assumed to be homogeneous, yet it consists of at least $\sim 10^{83}$ separate regions which are causally disconnected (i.e., these regions have not yet had time to communicate with each other via light signals).[5] (The precise assumptions which lead to these numbers will be spelled out in Sec. II.) Thus, one must assume that the forces which created these initial conditions were capable of violating causality.

The second puzzle is the flatness problem. This puzzle seems to be much less celebrated than the first, but it has been stressed by Dicke and Peebles.[6] I feel that it is of comparable importance to the first. It is known that the energy density ρ of the universe today is near the critical value ρ_{cr} (corresponding to the borderline between an open and closed universe). One can safely assume that[7]

$$0.01 < \Omega_p < 10, \tag{1.1}$$

where

$$\Omega \equiv \rho/\rho_{cr} = (8\pi/3)G\rho/H^2, \tag{1.2}$$

and the subscript p denotes the value at the present time. Although these bounds do not appear at first sight to be remarkably stringent, they, in fact, have powerful implications. The key point is that the condition $\Omega \approx 1$ is unstable. Furthermore, the only time scale which appears in the equations for a radiation-dominated universe is the Planck time, $1/M_P = 5.4\times 10^{-44}$ sec. A typical closed universe will reach its maximum size on the order of this time scale, while a typical open universe will dwindle to a value of ρ much less than ρ_{cr}. A universe can survive $\sim 10^{10}$ years only by extreme fine tuning of the initial values of ρ and H, so that ρ is very near ρ_{cr}. For the initial conditions taken at

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education,Bantalab Jammu

$T_0 = 10^{17}$ GeV, the value of H_0 must be fine tuned to an accuracy of one part in 10^{55}. In the standard model this incredibly precise initial relationship must be assumed without explanation. (For any reader who is not convinced that there is a real problem here, variations of this argument are given in the Appendix.)

The reader should not assume that these incredible numbers are due merely to the rather large value I have taken for T_0. If I had chosen a modest value such as $T_0 = 1$ MeV, I would still have concluded that the "initial" universe consisted of at least $\sim 10^{22}$ causally disconnected regions, and that the initial value of H_0 was fine tuned to one part in 10^{15}. These numbers are much smaller than the previous set, but they are still very impressive.

Of course, any problem involving the initial conditions can always be put off until we understand the physics of $T \gtrsim M_P$. However, it is the purpose of this paper to show that these puzzles might be obviated by a scenario for the behavior of the universe at temperatures well below M_P.

The paper is organized as follows. The assumptions and basic equations of the standard model are summarized in Sec. II. In Sec. III, I describe the inflationary universe scenario, showing how it can eliminate the horizon and flatness problems. The scenario is discussed in the context of grand models in Sec. IV, and comments are made concerning magnetic monopole suppression. In Sec. V I discuss briefly the key undesirable feature of the scenario: the inhomogeneities produced by the random nucleation of bubbles. Some vague ideas which might alleviate these difficulties are mentioned in Sec. VI.

II. THE STANDARD MODEL OF THE VERY EARLY UNIVERSE

In this section I will summarize the basic equations of the standard model, and I will spell out the assumptions which lead to the statements made in the Introduction.

The universe is assumed to be homogeneous and isotropic, and is therefore described by the Robertson-Walker metric[8]:

$$d\tau^2 = dt^2 - R^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right], \tag{2.1}$$

where $k = +1$, -1, or 0 for a closed, open, or flat universe, respectively. It should be emphasized that any value of k is possible, but by convention r and $R(t)$ are rescaled so that k takes on one of the three discrete values. The evolution of $R(t)$ is governed by the Einstein equations

$$\ddot{R} = -\frac{4\pi}{3} G(\rho + 3p)R, \tag{2.2a}$$

$$H^2 + \frac{k}{R^2} = \frac{8\pi}{3} G\rho, \tag{2.2b}$$

where $H \equiv \dot{R}/R$ is the Hubble "constant" (the dot denotes the derivative with respect to t). Conservation of energy is expressed by

$$\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}(R^3), \tag{2.3}$$

where p denotes the pressure. In the standard model one also assumes that the expansion is adiabatic, in which case

$$\frac{d}{dt}(sR^3) = 0, \tag{2.4}$$

where s is the entropy density.

To determine the evolution of the universe, the above equations must be supplemented by an equation of state for matter. It is now standard to describe matter by means of a field theory, and at high temperatures this means that the equation of state is to a good approximation that of an ideal quantum gas of massless particles. Let $N_b(T)$ denote the number of bosonic spin degrees of freedom which are effectively massless at temperature T (e.g., the photon contributes two units to N_b); and let $N_f(T)$ denote the corresponding number for fermions (e.g., electrons and positrons together contribute four units). Provided that T is not near any mass thresholds, the thermodynamic functions are given by

$$\rho = 3p = \frac{\pi^2}{30}\mathfrak{N}(T)T^4, \tag{2.5}$$

$$s = \frac{2\pi^2}{45}\mathfrak{N}(T)T^3, \tag{2.6}$$

$$n = \frac{\zeta(3)}{\pi^2}\mathfrak{N}'(T)T^3, \tag{2.7}$$

where

$$\mathfrak{N}(T) = N_b(T) + \tfrac{7}{8}N_f(T), \tag{2.8}$$

$$\mathfrak{N}'(T) = N_b(T) + \tfrac{3}{4}N_f(T). \tag{2.9}$$

Here n denotes the particle number density, and $\zeta(3) = 1.202\,06\ldots$ is the Riemann zeta function.

The evolution of the universe is then found by rewriting (2.2b) solely in terms of the temperature. Againing assuming that T is not near any mass thresholds, one finds

$$\left(\frac{\dot{T}}{T}\right)^2 + \epsilon(T)T^2 = \frac{4\pi^3}{45} G\mathfrak{N}(T)T^4, \tag{2.10}$$

where

$$\epsilon(T) = \frac{k}{R^2T^2} = k\left[\frac{2\pi^2}{45}\frac{\mathfrak{N}(T)}{S}\right]^{2/3}, \tag{2.11}$$

where $S \equiv R^3 s$ denotes the total entropy in a volume

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

specified by the radius of curvature R.

Since S is conserved, its value in the early universe can be determined (or at least bounded) by current observations. Taking $\rho < 10\rho_{cr}$ today, it follows that today

$$\left|\frac{k}{R^2}\right| < 9H^2. \tag{2.12}$$

From now on I will take $k = \pm 1$; the special case $k = 0$ is still included as the limit $R \to \infty$. Then today $R > \frac{1}{3}H^{-1} \sim 3\times 10^9$ years. Taking the present photon temperature T_γ as 2.7 °K, one then finds that the photon contribution to S is bounded by

$$S_\gamma > 3\times 10^{85}. \tag{2.13}$$

Assuming that there are three species of massless neutrinos (e, μ, and τ), all of which decouple at a time when the other effectively massless particles are the electrons and photons, then $S_\nu = 21/22 S_\gamma$. Thus,

$$S > 10^{86} \tag{2.14}$$

and

$$|\epsilon| < 10^{-58}\mathfrak{N}^{2/3}. \tag{2.15}$$

But then

$$\left|\frac{\rho-\rho_{cr}}{\rho}\right| = \frac{45}{4\pi^3}\frac{M_P^2}{\mathfrak{N}T^2}|\epsilon| < 3\times 10^{-59}\mathfrak{N}^{-1/3}(M_P/T)^2. \tag{2.16}$$

Taking $T = 10^{17}$ GeV and $\mathfrak{N} \sim 10^2$ (typical of grand unified models), one finds $|\rho-\rho_{cr}|/\rho < 10^{-55}$. This is the flatness problem.

The ϵT^2 term can now be deleted from (2.10), which is then solved (for temperatures higher than all particle masses) to give

$$T^2 = \frac{M_P}{2\gamma t}, \tag{2.17}$$

where $\gamma^2 = (4\pi^3/45)\mathfrak{N}$. (For the minimal SU_5 grand unified model, $N_b = 82$, $N_f = 90$, and $\gamma = 21.05$.) Conservation of entropy implies $RT =$ constant, so $R \propto t^{1/2}$. A light pulse beginning at $t = 0$ will have traveled by time t a physical distance

$$l(t) = R(t)\int_0^t dt' R^{-1}(t') = 2t, \tag{2.18}$$

and this gives the physical horizon distance. This horizon distance is to be compared with the radius $L(t)$ of the region at time t which will evolve into our currently observed region of the universe. Again using conservation of entropy,

$$L(t) = [s_p/s(t)]^{1/3}L_p, \tag{2.19}$$

where s_p is the present entropy density and $L_p \sim 10^{10}$ years is the radius of the currently observed region of the universe. One is interested in the ratio of volumes, so

$$\frac{l^3}{L^3} = \frac{11}{43}\left(\frac{45}{4\pi^3}\right)^{3/2}\mathfrak{N}^{-1/2}\left(\frac{M_P}{L_p T_\gamma T}\right)^3 = 4\times 10^{-89}\mathfrak{N}^{-1/2}(M_P/T)^3. \tag{2.20}$$

Taking $\mathfrak{N} \sim 10^2$ and $T_0 = 10^{17}$ GeV, one finds $l_0^3/L_0^3 = 10^{-83}$. This is the horizon problem.

III. THE INFLATIONARY UNIVERSE

In this section I will describe a scenario which is capable of avoiding the horizon and flatness problems.

From Sec. II one can see that both problems could disappear if the assumption of adiabaticity were grossly incorrect. Suppose instead that

$$S_p = Z^3 S_0, \tag{3.1}$$

where S_p and S_0 denote the present and initial values of $R^3 s$, and Z is some large factor.

Let us look first at the flatness problem. Given (3.1), the right-hand side (RHS) of (2.16) is multiplied by a factor of Z^2. The "initial" value (at $T_0 = 10^{17}$ GeV) of $|\rho-\rho_{cr}|/\rho$ could be of order unity, and the flatness problem would be obviated, if

$$Z > 3\times 10^{27}. \tag{3.2}$$

Now consider the horizon problem. The RHS of (2.19) is multiplied by Z^{-1}, which means that the length scale of the early universe, at any given temperature, was smaller by a factor of Z than had been previously thought. If Z is sufficiently large, then the initial region which evolved into our observed region of the universe would have been smaller than the horizon distance at that time. To see how large Z must be, note that the RHS of (2.20) is multiplied by Z^3. Thus, if

$$Z > 5\times 10^{27}, \tag{3.3}$$

then the horizon problem disappears. (It should be noted that the horizon will still exist; it will simply be moved out to distances which have not been observed.)

It is not surprising that the RHS's of (3.2) and (3.3) are approximately equal, since they both correspond roughly to S_0 of order unity.

I will now describe a scenario, which I call the inflationary universe, which is capable of such a large entropy production.

Suppose the equation of state for matter (with all chemical potentials set equal to zero) exhibits a first-order phase transition at some critical temperature T_c. Then as the universe cools through the temperature T_c, one would expect bubbles of the low-temperature phase to nucleate and grow. However, suppose the nucleation rate for this phase transition is rather low. The universe will

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

continue to cool as it expands, and it will then supercool in the high-temperature phase. Suppose that this supercooling continues down to some temperature T_s, many orders of magnitude below T_c. When the phase transition finally takes place at temperature T_s, the latent heat is released. However, this latent heat is characteristic of the energy scale T_c, which is huge relative to T_s. The universe is then reheated to some temperature T_r which is comparable to T_c. The entropy density is then increased by a factor of roughly $(T_r/T_s)^3$ (assuming that the number $\mathfrak{N}$ of degrees of freedom for the two phases are comparable), while the value of R remains unchanged. Thus,

$$Z \approx T_r/T_s . \tag{3.4}$$

If the universe supercools by 28 or more orders of magnitude below some critical temperature, the horizon and flatness problems disappear.

In order for this scenario to work, it is necessary for the universe to be essentially devoid of any strictly conserved quantities. Let n denote the density of some strictly conserved quantity, and let $r \equiv n/s$ denote the ratio of this conserved quantity to entropy. Then $r_f = Z^{-3} r_0 < 10^{-84} r_0$. Thus, only an absurdly large value for the initial ratio would lead to a measurable value for the present ratio. Thus, if baryon number were exactly conserved, the inflationary model would be untenable. However, in the context of grand unified models, baryon number is not exactly conserved. The net baryon number of the universe is believed to be created by CP-violating interactions at a temperature of 10^{13}–10^{14} GeV.[9] Thus, provided that T_c lies in this range or higher, there is no problem. The baryon production would take place after the reheating. (However, strong constraints are imposed on the entropy which can be generated in any phase transition with $T_c \ll 10^{14}$ GeV, in particular, the Weinberg-Salam phase transition.[36])

Let us examine the properties of the supercooling universe in more detail. Note that the energy density $\rho(T)$, given in the standard model by (2.5), must now be modified. As $T \to 0$, the system is cooling not toward the true vacuum, but rather toward some metastable false vacuum with an energy density ρ_0 which is necessarily higher than that of the true vacuum. Thus, to a good approximation (ignoring mass thresholds)

$$\rho(T) = \frac{\pi^2}{30}\mathfrak{N}(T)T^4 + \rho_0 . \tag{3.5}$$

Perhaps a few words should be said concerning the zero point of energy. Classical general relativity couples to an energy-momentum tensor of matter, $T_{\mu\nu}$, which is necessarily (covariantly) conserved. When matter is described by a field theory, the form of $T_{\mu\nu}$ is determined by the conservation requirement up to the possible modification

$$T_{\mu\nu} \to T_{\mu\nu} + \lambda g_{\mu\nu} , \tag{3.6}$$

for any constant λ. (λ *cannot* depend on the values of the fields, nor can it depend on the temperature or the phase.) The freedom to introduce the modification (3.6) is identical to the freedom to introduce a cosmological constant into Einstein's equations. One can always choose to write Einstein's equations without an explicit cosmological term; the cosmological constant Λ is then defined by

$$\langle 0 | T_{\mu\nu} | 0 \rangle = \Lambda g_{\mu\nu} , \tag{3.7}$$

where $|0\rangle$ denotes the true vacuum. Λ is identified as the energy density of the vacuum, and, in principle, there is no reason for it to vanish. Empirically Λ is known to be very small ($|\Lambda| < 10^{-46}$ GeV^4)[10] so I will take its value to be zero.[11] The value of ρ_0 is then necessarily positive and is determined by the particle theory.[12] It is typically of $O(T_c^4)$.

Using (3.5), Eq. (2.10) becomes

$$\left(\frac{\dot{T}}{T}\right)^2 = \frac{4\pi^3}{45} G\mathfrak{N}(T)T^4 - \epsilon(T)T^2 + \frac{8\pi}{3} G\rho_0 . \tag{3.8}$$

This equation has two types of solutions, depending on the parameters. If $\epsilon > \epsilon_0$, where

$$\epsilon_0 = \frac{8\pi^2\sqrt{30}}{45} G\sqrt{\mathfrak{N}\rho_0} , \tag{3.9}$$

then the expansion of the universe is halted at a temperature T_{min} given by

$$T_{min}^4 = \frac{30\rho_0}{\pi^2}\left\{\frac{\epsilon}{\epsilon_0} + \left[\left(\frac{\epsilon}{\epsilon_0}\right)^2 - 1\right]^{1/2}\right\}^2 , \tag{3.10}$$

and then the universe contracts again. Note that T_{min} is of $O(T_c)$, so this is not the desired scenario. The case of interest is $\epsilon < \epsilon_0$, in which case the expansion of the universe is unchecked. [Note that $\epsilon_0 \sim \sqrt{\mathfrak{N}} T_c^2/M_P$ is presumably a very small number. Thus $0 < \epsilon < \epsilon_0$ (a closed universe) seems unlikely, but $\epsilon < 0$ (an open universe) is quite plausible.] Once the temperature is low enough for the ρ_0 term to dominate over the other two terms on the RHS of (3.8), one has

$$T(t) \approx \text{const} \times e^{-\chi t} , \tag{3.11}$$

where

$$\chi^2 = \frac{8\pi}{3} G\rho_0 . \tag{3.12}$$

Since $RT = \text{const}$, one has[13]

$$R(t) = \text{const} \times e^{\chi t} . \tag{3.13}$$

The universe is expanding exponentially, in a false

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

vacuum state of energy density ρ_0. The Hubble constant is given by $H=\dot{R}/R=\chi$. (More precisely, H approaches χ monotonically from above. This behavior differs markedly from the standard model, in which H falls as t^{-1}.)

The false vacuum state is Lorentz invariant, so $T_{\mu\nu}=\rho_0 g_{\mu\nu}$. It follows that $p=-\rho_0$, the pressure is negative. This negative pressure allows for the conservation of energy, Eq. (2.3). From the second-order Einstein equation (2.2a), it can be seen that the negative pressure is also the driving force behind the exponential expansion.

The Lorentz invariance of the false vacuum has one other consequence: The metric described by (3.13) (with $k=0$) does not single out a comoving frame. The metric is invariant under an O(4,1) group of transformations, in contrast to the usual Robertson-Walker invariance of O(4).[14] It is known as the de Sitter metric, and it is discussed in the standard literature.[15]

Now consider the process of bubble formation in a Robertson-Walker universe. The bubbles form randomly, so there is a certain nucleation rate $\lambda(t)$, which is the probability per (physical) volume per time that a bubble will form in any region which is still in the high-temperature phase. I will idealize the situation slightly and assume that the bubbles start at a point and expand at the speed of light. Furthermore, I neglect k in the metric, so $d\tau^2=dt^2-R^2(t)d\vec{x}^2$.

I want to calculate $p(t)$, the probability that any given point remains in the high-temperature phase at time t. Note that the distribution of bubbles is totally uncorrelated except for the exclusion principle that bubbles do not form inside of bubbles. This exclusion principle causes no problem because one can imagine fictitious bubbles which form inside the real bubbles with the same nucleation rate $\lambda(t)$. With all bubbles expanding at the speed of light, the fictitious bubbles will be forever inside the real bubbles and will have no effect on $p(t)$. The distribution of all bubbles, real and fictitious, is then totally uncorrelated.

$p(t)$ is the probability that there are no bubbles which engulf a given point in space. But the number of bubbles which engulf a given point is a Poisson-distributed variable, so $p(t)=\exp[-\bar{N}(t)]$, where $\bar{N}(t)$ is the expectation value of the number of bubbles engulfing the point. Thus[16]

$$p(t)=\exp\left[-\int_0^t dt_1\lambda(t_1)R^3(t_1)V(t,t_1)\right], \tag{3.14}$$

where

$$V(t,t_1)=\frac{4\pi}{3}\left[\int_{t_1}^{t}\frac{dt_2}{R(t_2)}\right]^3 \tag{3.15}$$

is the coordinate volume at time t of a bubble which formed at time t_1.

I will now assume that the nucleation rate is sufficiently slow so that no significant nucleation takes place until $T\ll T_c$, when exponential growth has set in. I will further assume that by this time $\lambda(t)$ is given approximately by the zero-temperature nucleation rate λ_0. One then has

$$p(t)=\exp\left[-\frac{t}{\tau}+O(1)\right], \tag{3.16}$$

where

$$\tau=\frac{3\chi^3}{4\pi\lambda_0}, \tag{3.17}$$

and $O(1)$ refers to terms which approach a constant as $\chi t\to\infty$. During one of these time constants, the universe will expand by a factor

$$Z_\tau=\exp(\chi\tau)=\exp\left(\frac{3\chi^4}{4\pi\lambda_0}\right). \tag{3.18}$$

If the phase transition is associated with the expectation value of a Higgs field, then λ_0 can be calculated using the method of Coleman and Callan.[17] The key point is that nucleation is a tunneling process, so that λ_0 is typically very small. The Coleman-Callan method gives an answer of the form

$$\lambda_0=A\rho_0\exp(-B), \tag{3.19}$$

where B is a barrier penetration term and A is a dimensionless coefficient of order unity. Since Z_τ is then an exponential of an exponential, one can very easily[18,19,36] obtain values as large as $\log_{10}Z\approx 28$, or even $\log_{10}Z\approx 10^{10}$.

Thus, if the universe reaches a state of exponential growth, it is quite plausible for it to expand and supercool by a huge number of orders of magnitude before a significant fraction of the universe undergoes the phase transition.

So far I have assumed that the early universe can be described from the beginning by a Robertson-Walker metric. If this assumption were really necessary, then it would be senseless to talk about "solving" the horizon problem; perfect homogeneity was assumed at the outset. Thus, I must now argue that the assumption can probably be dropped.

Suppose instead that the initial metric, and the distribution of particles, was rather chaotic. One would then expect that statistical effects would tend to thermalize the particle distribution on a local scale.[20] It has also been shown (in idealized circumstances) that anisotropies in the metric are damped out on the time scale of $\sim 10^3$ Planck times.[21] The damping of inhomogeneities in the metric has also been studied,[22] and it is reasonable to expect such damping to occur. Thus, assuming that at least some region of the universe started at

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

temperatures high compared to T_c, one would expect that, by the time the temperature in one of these regions falls to T_c, it will be *locally* homogeneous, isotropic, and in thermal equilibrium. By locally, I am talking about a length scale ξ which is of course less than the horizon distance. It will then be possible to describe this local region of the universe by a Robertson-Walker metric, which will be accurate at distance scales small compared to ξ. When the temperature of such a region falls below T_c, the inflationary scenario will take place. The end result will be a huge region of space which is homogeneous, isotropic, and of nearly critical mass density. If Z is sufficiently large, this region can be bigger than (or much bigger than) our observed region of the universe.

IV. GRAND UNIFIED MODELS AND MAGNETIC MONOPOLE PRODUCTION

In this section I will discuss the inflationary model in the context of grand unified models of elementary-particle interactions.[23,24]

A grand unified model begins with a simple gauge group G which is a valid symmetry at the highest energies. As the energy is lowered, the theory undergoes a hierarchy of spontaneous symmetry breaking into successive subgroups: $G \to H_n \to \cdots \to H_0$, where $H_1 = SU_3 \times SU_2 \times U_1$ [QCD (quantum chromodynamics) × Weinberg-Salam] and $H_0 = SU_3 \times U_1^{EM}$. In the Georgi-Glashow model,[23] which is the simplest model of this type, $G = SU_5$ and $n = 1$. The symmetry breaking of $SU_5 \to SU_3 \times SU_2 \times U_1$ occurs at an energy scale $M_X \sim 10^{14}$ GeV.

At high temperatures, it was suggested by Kirzhnits and Linde[25] that the Higgs fields of any spontaneously broken gauge theory would lose their expectation values, resulting in a high-temperature phase in which the full gauge symmetry is restored. A formalism for treating such problems was developed[26] by Weinberg and by Dolan and Jackiw. In the range of parameters for which the tree potential is valid, the phase structure of the SU_5 model was analyzed by Tye and me.[16,27] We found that the SU_5 symmetry is restored at $T > \sim 10^{14}$ GeV and that for most values of the parameters there is an intermediate-temperature phase with gauge symmetry $SU_4 \times U_1$, which disappears at $T \sim 10^{13}$ GeV. Thus, grand unified models tend to provide phase transitions which could lead to an inflationary scenario of the universe.

Grand unified models have another feature with important cosmological consequences: They contain very heavy magnetic monopoles in their particle spectrum. These monopoles are of the type discovered by 't Hooft and Polyakov,[28] and will be present in any model satisfying the above description.[29] These monopoles typically have masses of order $M_X/\alpha \sim 10^{16}$ GeV, where $\alpha = g^2/4\pi$ is the grand unified fine structure constant. Since the monopoles are really topologically stable knots in the Higgs field expectation value, they do not exist in the high-temperature phase of the theory. They therefore come into existence during the course of a phase transition, and the dynamics of the phase transition is then intimately related to the monopole production rate.

The problem of monopole production and the subsequent annihilation of monopoles, in the context of a second-order or weakly first-order phase transition, was analyzed by Zeldovich and Khlopov[30] and by Preskill.[31] In Preskill's analysis, which was more specifically geared toward grand unified models, it was found that relic monopoles would exceed present bounds by roughly 14 orders of magnitude. Since it seems difficult to modify the estimated annihilation rate, one must find a scenario which suppresses the production of these monopoles.

Kibble[32] has pointed out that monopoles are produced in the course of the phase transition by the process of bubble coalescence. The orientation of the Higgs field inside one bubble will have no correlation with that of another bubble not in contact. When the bubbles coalesce to fill the space, it will be impossible for the uncorrelated Higgs fields to align uniformly. One expects to find topological knots, and these knots are the monopoles. The number of monopoles so produced is then comparable to the number of bubbles, to within a few orders of magnitude.

Kibble's production mechanism can be used to set a "horizon bound" on monopole production which is valid if the phase transition does not significantly disturb the evolution of the universe.[33] At the time of bubble coalescence t_{coal} the size l of the bubbles cannot exceed the horizon distance at that time. So

$$l < 2t_{coal} = \frac{M_P}{\gamma T_{coal}^2} . \tag{4.1}$$

By Kibble's argument, the density n_M of monopoles then obeys

$$n_M \geq l^{-3} > \frac{\gamma^3 T_{coal}^6}{M_P^3} . \tag{4.2}$$

By considering the contribution to the mass density of the present universe which could come from 10^{16} GeV monopoles, Preskill[31] concludes that

$$n_M/n_\gamma < 10^{-24} , \tag{4.3}$$

where n_γ is the density of photons. This ratio changes very little from the time of the phase transition, so with (2.7) one concludes

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$T_{\rm coal} < \left[\frac{10^{-24}\pi^2}{2\zeta(3)}\right]^{1/3} \gamma^{-1} M_P \approx 10^{10}\ {\rm GeV}\,. \qquad (4.4)$$

If $T_c \sim 10^{14}$ GeV, this bound implies that the universe must supercool by at least about four orders of magnitude before the phase transition is completed.

The problem of monopole production in a strongly first-order phase transition with supercooling was treated in more detail by Tye and me.[16,34] We showed how to explicitly calculate the bubble density in terms of the nucleation rate, and we considered the effects of the latent heat released in the phase transition. Our conclusion was that (4.4) should be replaced by

$$T_{\rm coal} < 2\times 10^{11}\ {\rm GeV}\,, \qquad (4.5)$$

where $T_{\rm coal}$ refers to the temperature just before the release of the latent heat.

Tye and I omitted the crucial effects of the mass density ρ_0 of the false vacuum. However, our work has one clear implication: If the nucleation rate is sufficiently large to avoid exponential growth, then far too many monopoles would be produced. Thus, the monopole problem seems to also force one into the inflationary scenario.[35]

In the simplest SU_5 model, the nucleation rates have been calculated (approximately) by Weinberg and me.[19] The model contains unknown parameters, so no definitive answer is possible. We do find, however, that there is a sizable range of parameters which lead to the inflationary scenario.[36]

V. PROBLEMS OF THE INFLATIONARY SCENARIO[37]

As I mentioned earlier, the inflationary scenario seems to lead to some unacceptable consequences. It is hoped that some variation can be found which avoids these undesirable features but maintains the desirable ones. The problems of the model will be discussed in more detail elsewhere,[37] but for completeness I will give a brief description here.

The central problem is the difficulty in finding a smooth ending to the period of exponential expansion. Let us assume that $\lambda(t)$ approaches a constant as $t \to -\infty$ and $T \to 0$. To achieve the desired expansion factor $Z > 10^{28}$, one needs $\lambda_0/\chi^4 < 10^{-2}$ [see (3.18)], which means that the nucleation rate is slow compared to the expansion rate of the universe. (Explicit calculations show that λ_0/χ^4 is typically much smaller than this value.[18,19,36]) The randomness of the bubble formation process then leads to gross inhomogeneities.

To understand the effects of this randomness, the reader should bear in mind the following facts.

(i) All of the latent heat released as a bubble expands is transferred initially to the walls of the bubble.[17] This energy can be thermalized only when the bubble walls undergo many collisions.

(ii) The de Sitter metric does not single out a comoving frame. The O(4, 1) invariance of the de Sitter metric is maintained even after the formation of one bubble. The memory of the original Robertson-Walker comoving frame is maintained by the probability distribution of bubbles, but the local comoving frame can be reestablished only after enough bubbles have collided.

(iii) The size of the largest bubbles will exceed that of the smallest bubbles by roughly a factor of Z; the range of bubble sizes is immense. The surface energy density grows with the size of the bubble, so the energy in the walls of the largest bubbles can be thermalized only by colliding with other large bubbles.

(iv) As time goes on, an arbitrarily large fraction of the space will be in the new phase [see (3.16)]. However, one can ask a more subtle question about the region of space which is in the new phase: Is the region composed of finite separated clusters, or do these clusters join together to form an infinite region? The latter possibility is called "percolation." It can be shown[38] that the system percolates for large values of λ_0/χ^4, but that for sufficiently small values it does *not*. The critical value of λ_0/χ^4 has not been determined, but presumably an inflationary universe would have a value of λ_0/χ^4 below critical. Thus, no matter how long one waits, the region of space in the new phase will consist of finite clusters, each totally surrounded by a region in the old phase.

(v) Each cluster will contain only a few of the largest bubbles. Thus, the collisions discussed in (iii) cannot occur.

The above statements do not quite prove that the scenario is impossible, but these consequences are at best very unattractive. Thus, it seems that the scenario will become viable only if some modification can be found which avoids these inhomogeneities. Some vague possibilities will be mentioned in the next section.

Note that the above arguments seem to rule out the possibility that the universe was ever trapped in a false vacuum state, unless $\lambda_0/\chi^4 \gtrsim 1$. Such a large value of λ_0/χ^4 does not seem likely, but it is possible.[19]

VI. CONCLUSION

I have tried to convince the reader that the standard model of the very early universe requires the assumption of initial conditions which are very implausible for two reasons:

(i) *The horizon problem.* Causally disconnected regions are assumed to be nearly identical; in par-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ticular, they are simultaneously at the same temperature.

(ii) *The flatness problem.* For a fixed initial temperature, the initial value of the Hubble "constant" must be fine tuned to extraordinary accuracy to produce a universe which is as flat as the one we observe.

Both of these problems would disappear if the universe supercooled by 28 or more orders of magnitude below the critical temperature for some phase transition. (Under such circumstances, the universe would be growing exponentially in time.) However, the random formation of bubbles of the new phase seems to lead to a much too inhomogeneous universe.

The inhomogeneity problem would be solved if one could avoid the assumption that the nucleation rate $\lambda(t)$ approaches a small constant λ_0 as the temperature $T \to 0$. If, instead, the nucleation rate rose sharply at some T_1, then bubbles of an approximately uniform size would suddenly fill space as T fell to T_1. Of course, the full advantage of the inflationary scenario is achieved only if $T_1 \lesssim 10^{-28} T_c$.

Recently Witten[39] has suggested that the above chain of events may in fact occur if the parameters of the SU_5 Higgs field potential are chosen to obey the Coleman-Weinberg condition[40] (i.e., that $\partial^2 V/\partial\phi^2 = 0$ at $\phi = 0$). Witten[41] has studied this possibility in detail for the case of the Weinberg-Salam phase transition. Here he finds that thermal tunneling is totally ineffective, but instead the phase transition is driven when the temperature of the QCD chiral-symmetry-breaking phase transition is reached. For the SU_5 case, one can hope that a much larger amount of supercooling will be found; however, it is difficult to see how 28 orders of magnitude could arise.

Another physical effect which has so far been left out of the analysis is the production of particles due to the changing gravitational metric.[42] This effect may become important in an exponentially expanding universe at low temperatures.

In conclusion, the inflationary scenario seems like a natural and simple way to eliminate both the horizon and the flatness problems. I am publishing this paper in the hope that it will highlight the existence of these problems and encourage others to find some way to avoid the undesirable features of the inflationary scenario.

ACKNOWLEDGMENTS

I would like to express my thanks for the advice and encouragement I received from Sidney Coleman and Leonard Susskind, and for the invaluable help I received from my collaborators Henry Tye and Erick Weinberg. I would also like to acknowledge very useful conversations with Michael Aizenman, Beilok Hu, Harry Kesten, Paul Langacker, Gordon Lasher, So-Young Pi, John Preskill, and Edward Witten. This work was supported by the Department of Energy under Contract No. DE-AC03-76SF00515.

APPENDIX: REMARKS ON THE FLATNESS PROBLEM

This appendix is added in the hope that some skeptics can be convinced that the flatness problem is real. Some physicists would rebut the argument given in Sec. I by insisting that the equations might make sense all the way back to $t = 0$. Then if one fixes the value of H corresponding to some arbitrary temperature T_a, one always finds that when the equations are extrapolated *backward* in time, $\Omega \to 1$ as $t \to 0$. Thus, they would argue, it is natural for Ω to be very nearly equal to 1 at early times. For physicists who take this point of view, the flatness problem must be restated in other terms. Since H_0 and T_0 have no significance, the model universe must be specified by its conserved quantities. In fact, the model universe is completely specified by the dimensionless constant $\epsilon \equiv k/R^2T^2$, where k and R are parameters of the Robertson-Walker metric, Eq. (2.1). For our universe, one must take $|\epsilon| < 3 \times 10^{-57}$. The problem then is the to explain why $|\epsilon|$ should have such a startlingly small value.

Some physicists also take the point of view that $\epsilon \equiv 0$ is plausible enough, so to them there is no problem. To these physicists I point out that the universe is certainly not described *exactly* by a Robertson-Walker metric. Thus it is difficult to imagine any physical principle which would require a parameter of that metric to be exactly equal to zero.

In the end, I must admit that questions of plausibility are not logically determinable and depend somewhat on intuition. Thus I am sure that some physicists will remain unconvinced that there really is a flatness problem. However, I am also sure that many physicists agree with me that the flatness of the universe is a peculiar situation which at some point will admit a physical explanation.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

*Present address: Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

[1]I use units for which $\hbar = c = k$ (Boltzmann constant) = 1. Then 1 m = 5.068×10^{15} GeV^{-1}, 1 kg = 5.610×10^{26} GeV, 1 sec = 1.519×10^{24} GeV^{-1}, and 1 °K = 8.617×10^{-14} GeV.

[2]W. Rindler, Mon. Not. R. Astron. Soc. 116, 663 (1956). See also Ref. 3, pp. 489–490, 525–526; and Ref. 4, pp. 740 and 815.

[3]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).

[4]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[5]In order to calculate the horizon distance, one must of course follow the light trajectories back to $t = 0$. This violates my contention that the equations are to be trusted only for $T \lesssim T_0$. Thus, the horizon problem could be obviated if the full quantum gravitational theory had a radically different behavior from the naive extrapolation. Indeed, solutions of this sort have been proposed by A. Zee, Phys. Rev. Lett. 44, 703 (1980) and by F. W. Stecker, Astrophys. J. 235, L1 (1980). However, it is the point of this paper to show that the horizon problem can also be obviated by mechanisms which are more within our grasp, occurring at temperatures below T_0.

[6]R. H. Dicke and P. J. E. Peebles, *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, London, 1979).

[7]See Ref. 3, pp. 475–481; and Ref. 4, pp. 796–797.

[8]For example, see Ref. 3, Chap. 14.

[9]M. Yoshimura, Phys. Rev. Lett. 41, 281 (1978); 42, 746(E) (1979); Phys. Lett. 88B, 294 (1979); S. Dimopoulos and L. Susskind, Phys. Rev. D 18, 4500 (1978); Phys. Lett. 81B, 416 (1979); A. Yu Ignatiev, N. V. Krashikov, V. A. Kuzmin, and A. N. Tavkhelidze, *ibid.* 76B, 436 (1978); D. Toussaint, S. B. Treiman, F. Wilczek, and A. Zee, Phys. Rev. D 19, 1036 (1979); S. Weinberg, Phys. Rev. Lett. 42, 850 (1979); D. V. Nanopoulos and S. Weinberg, Phys. Rev. D 20, 2484 (1979); J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, Phys. Lett. 80B, 360 (1979); 82B, 464 (1979); M. Honda and M. Yoshimura, Prog. Theor. Phys. 62, 1704 (1979); D. Toussaint and F. Wilczek, Phys. Lett. 81B, 238 (1979); S. Barr, G. Segre, and H. A. Weldon, Phys. Rev. D 20, 2494 (1979); A. D. Sakharov, Zh. Eksp. Teor. Fiz. 76, 1172 (1979) [Sov. Phys.—JETP 49, 594 (1979)]; A. Yu Ignatiev, N. V. Krashikov, V. A. Kuzmin, and M. E. Shaposhnikov, Phys. Lett. 87B, 114 (1979); E. W. Kolb and S. Wolfram, *ibid.* 91B, 217 (1980); Nucl. Phys. B172, 224 (1980); J. N. Fry, K. A. Olive, and M. S. Turner, Phys. Rev. D 22, 2953 (1980); 22, 2977 (1980); S. B. Treiman and F. Wilczek, Phys. Lett. 95B, 222 (1980); G. Senjanović and F. W. Stecker, Phys. Lett. B (to be published).

[11]The reason Λ is so small is of course one of the deep mysteries of physics. The value of Λ is not determined by the particle theory alone, but must be fixed by whatever theory couples particles to quantum gravity. This appears to be a separate problem from the ones discussed in this paper, and I merely use the empirical fact that $\Lambda \approx 0$.

[12]S. A. Bludman and M. A. Ruderman, Phys. Rev. Lett. 38, 255 (1977).

[13]The effects of a false vacuum energy density on the evolution of the early universe have also been considered by E. W. Kolb and S. Wolfram, CAL TECH Report No. 79-0984 (unpublished), and by S. A. Bludman, University of Pennsylvania Report No. UPR-0143T, 1979 (unpublished).

[14]More precisely, the usual invariance is O(4) if $k = 1$, O(3,1) if $k = -1$, and the group of rotations and translations in three dimensions if $k = 0$.

[15]See for example, Ref. 3, pp. 385–392.

[16]A. H. Guth and S.-H. Tye, Phys. Rev. Lett. 44, 631 (1980); 44, 963 (1980).

[17]S. Coleman, Phys. Rev. D 15, 2929 (1977); C. G. Callan and S. Coleman, *ibid.* 16, 1762 (1977); see also S. Coleman, in *The Whys of Subnuclear Physics*, proceedings of the International School of Subnuclear Physics, Ettore Majorana, Erice, 1977, edited by A. Zichichi (Plenum, New York, 1979).

[18]A. H. Guth and E. J. Weinberg, Phys. Rev. Lett. 45, 1131 (1980).

[19]E. J. Weinberg and I are preparing a manuscript on the possible cosmological implications of the phase transitions in the SU_5 grand unified model.

[20]J. Ellis and G. Steigman, Phys. Lett. 89B, 186 (1980); J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, *ibid.* 90B, 253 (1980).

[21]B. L. Hu and L. Parker, Phys. Rev. D 17, 933 (1978).

[22]See Ref. 4, Chap. 30.

[23]The simplest grand unified model is the SU(5) model of H. Georgi and S. L. Glashow, Phys. Rev. Lett. 32, 438 (1974). See also H. Georgi, H. R. Quinn, and S. Weinberg, *ibid.* 33, 451 (1974); and A. J. Buras, J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, Nucl. Phys. B135, 66 (1978).

[24]Other grand unified models include the SO(10) model: H. Georgi, in *Particles and Fields—1975*, proceedings of the 1975 meeting of the Division of Particles and Fields of the American Physical Society, edited by Carl Carlson (AIP, New York, 1975); H. Fritzsch and P. Minkowski, Ann. Phys. (N.Y.) 93, 193 (1975); H. Georgi and D. V. Nanopoulos, Phys. Lett. 82B, 392 (1979) and Nucl. Phys. B155, 52 (1979). The E(6) model: F. Gürsey, P. Ramond, and P. Sikivie, Phys. Lett. 60B, 177 (1975); F. Gürsey and M. Serdaroglu, Lett. Nuovo Cimento 21, 28 (1978). The E(7) model: F. Gürsey and P. Sikivie, Phys. Rev. Lett. 36, 775 (1976), and Phys. Rev. D 16, 816 (1977); P. Ramond, Nucl. Phys. B110, 214 (1976). For some general properties of grand unified models, see M. Gell-Mann, P. Ramond, and R. Slansky, Rev. Mod. Phys. 50, 721 (1978). For a review, see P. Langacker, Report No. SLAC-PUB-2544, 1980 (unpublished).

[25]D. A. Kirzhnits and A. D. Linde, Phys. Lett. 42B, 471 (1972).

[26]S. Weinberg, Phys. Rev. D 9, 3357 (1974); L. Dolan and R. Jackiw, *ibid.* 9, 3320 (1974); see also D. A. Kirzhnits and A. D. Linde, Ann. Phys. (N.Y.) 101, 195 (1976); A. D. Linde, Rep. Prog. Phys. 42, 389 (1979). ϵ-expansion techniques are employed by P. Ginsparg, Nucl. Phys. B (to be published).

[27]In the case that the Higgs quartic couplings are comparable to g^4 or smaller (g = gauge coupling), the phase structure has been studied by M. Daniel and C. E. Vayonakis, CERN Report No. TH.2860 1980

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

(unpublished); and by P. Suranyi, University of Cincinnati Report No. 80-0506 (unpublished).
[28]G. 't Hooft, Nucl. Phys. B79, 276 (1974); A. M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. 20, 430 (1974) [JETP Lett. 20, 194 (1974)]. For a review, see P. Goddard and D. I. Olive, Rep. Prog. Phys. 41, 1357 (1978).
[29]If $\Pi_1(G)$ and $\Pi_2(G)$ are both trivial, then $\Pi_2(G/H_0) = \Pi_1(H_0)$. In our case $\Pi_1(H_0)$ is the group of integers. For a general review of topology written for physicists, see N. D. Mermin, Rev. Mod. Phys. 51, 591 (1979).
[30]Y. B. Zeldovich and M. Y. Khlopov, Phys. Lett. 79B, 239 (1978).
[31]J. P. Preskill, Phys. Rev. Lett. 43, 1365 (1979).
[32]T. W. B. Kibble, J. Phys. A 9, 1387 (1976).
[33]This argument was first shown to me by John Preskill. It is also described by Einhorn *et al.*, Ref. 34, except that they make no distinction between T_{coal} and T_c.
[34]The problem of monopole production was also examined by M. B. Einhorn, D. L. Stein, and D. Toussaint, Phys. Rev. D 21, 3295 (1980), who focused on second-order transitions. The structure of SU(5) monopoles has been studied by C. P. Dokos and T. N. Tomaras, Phys. Rev. D 21, 2940 (1980); and by M. Daniel, G. Lazarides, and Q. Shafi, Nucl. Phys. B170, 156 (1980). The problem of suppression of the cosmological production of monopoles is discussed by G. Lazarides and Q. Shafi, Phys. Lett. 94B, 149 (1980), and G. Lazarides, M. Magg, and Q. Shafi, CERN Report No. TH.2856, 1980 (unpublished); the suppression discussed here relies on a novel confinement mechanism, and also on the same kind of supercooling as in Ref. 16. See also J. N. Fry and D. N. Schramm, Phys. Rev. Lett. 44, 1361 (1980).
[35]An alternative solution to the monopole problem has been proposed by P. Langacker and S.-Y. Pi, Phys. Rev. Lett. 45, 1 (1980). By modifying the Higgs structure, they have constructed a model in which the high-temperature SU_5 phase undergoes a phase transition to an SU_3 phase at $T \sim 10^{14}$GeV. Another phase transition occurs at $T \sim 10^3$ GeV, and below this temperature the symmetry is $SU_3 \times U_1^{EM}$. Monopoles cannot exist until $T < 10^3$ GeV, but their production is negligible at these low temperatures. The suppression of monopoles due to the breaking of U_1^{EM} symmetry at high temperatures was also suggested by S.-H. Tye, talk given at the 1980 Guangzhou Conference on Theoretical Particle Physics, Canton, 1980 (unpublished).
[36]The Weinberg-Salam phase transition has also been investigated by a number of authors: E. Witten, Ref. 41; M. A. Sher, Phys. Rev. D 22, 2989 (1980); P. J. Steinhardt, Harvard report, 1980 (unpublished); and A. H. Guth and E. J. Weinberg, Ref. 18.
[37]This section represents the work of E. J. Weinberg, H. Kesten, and myself. Weinberg and I are preparing a manuscript on this subject.
[38]The proof of this statement was outlined by H. Kesten (Dept. of Mathematics, Cornell University), with details completed by me.
[39]E. Witten, private communication.
[40]S. Coleman and E. J. Weinberg, Phys. Rev. D 7, 1888 (1973); see also, J. Ellis, M. K. Gaillard, D. Nanopoulos, and C. Sachrajda, Phys. Lett. 83B, 339 (1979), and J. Ellis, M. K. Gaillard, A. Peterman, and C. Sachrajda, Nucl. Phys. B164, 253 (1980).
[41]E. Witten, Nucl. Phys. B (to be published).
[42]L. Parker, in *Asymptotic Structure of Spacetime*, edited by F. Esposito and L. Witten (Plenum, New York, 1977); V. N. Lukash, I. D. Novikov, A. A. Starobinsky, and Ya. B. Zeldovich, Nuovo Cimento 35B, 293 (1976).

Erratum

Eq. (4.4) should read

$$T_{coal} < \left[10^{-24}\,\frac{2\zeta(3)}{\pi^2}\right]^{1/3} \gamma^{-1} M_p \approx 10^{10}\ \text{GeV}$$

Addenda

Since the publication of this article, many of the papers which were cited as preprints have been published:

Ref. 9: G. Senjanovic and F. W. Stecker, *Phys. Lett.* **96B**, 285 (1980).
Ref. 19: A. H. Guth and E. J. Weinberg, *Phys. Rev.* **D23**, 876 (1981).
Ref. 24: P. Langacker, *Phys. Reports* **72C**, 185 (1981).
Ref. 26: P. Ginsparg, *Nucl. Phys.* **B170** [FS1], 388 (1980).
Ref. 27: M. Daniel and C. E. Vayonakis, *Nucl. Phys.* **B180**, 301 (1981).
Ref. 34: G. Lazarides, M. Magg, and Q. Shafi, *Phys. Lett.* **97B**, 87 (1980).
Ref. 36: P. J. Steinhardt, *Nucl. Phys.* **B179**, 492 (1981).
Ref. 37: A. H. Guth and E. J. Weinberg, *Nucl. Phys.* **B212**, 321 (1983).
Ref. 41: E. Witten, *Nucl. Phys.* **B177**, 477 (1981).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Volume 108B, number 6 PHYSICS LETTERS 4 February 1982

A NEW INFLATIONARY UNIVERSE SCENARIO: A POSSIBLE SOLUTION OF THE HORIZON, FLATNESS, HOMOGENEITY, ISOTROPY AND PRIMORDIAL MONOPOLE PROBLEMS

A.D. LINDE
Lebedev Physical Institute, Moscow 117924, USSR

Received 29 October 1981

A new inflationary universe scenario is suggested, which is free of the shortcomings of the previous one and provides a possible solution of the horizon, flatness, homogeneity and isotropy problems in cosmology, and also a solution of the primordial monopole problem in grand unified theories.

There is now considerable interest in the cosmological consequences of symmetry breaking phase transitions, which occur in grand unified theories (GUTs) with the decrease of temperature at the very early stages of the evolution of the universe [1–3]. These phase transitions typically are strongly first order [4,5]. The lifetime of the supercooled symmetric phase $\varphi = 0$ (φ is the Higgs scalar field which breaks the symmetry) in some theories may be extremely large [2,3,6–8]. In that case the energy–momentum tensor of particles $\sim T^4$ in the phase $\varphi = 0$ almost vanishes in the course of the expansion of the universe, and the total energy–momentum tensor reduces to the vacuum stress tensor (cosmological term) $T^{\rm vac}_{\mu\nu} = g_{\mu\nu} V(0)$, where $V(\varphi)$ is the effective potential of the theory at vanishing temperature [2,3]. This leads to an exponentially fast expansion of the universe, $a \sim e^{Ht}$. Here a is the scale factor, and H is the Hubble constant at that time,

$$H = [(8\pi/3M_{\rm P}^2) V(0)]^{1/2},$$

where $M_{\rm P} \approx 10^{19}$ GeV is the Planck mass [9]. Then at some comparatively small temperature $T_{\rm c}$ the symmetry breaking phase transition takes place, all the vacuum energy $V_{(0)}$ transforms into thermal energy [2,3], the universe is reheated up to the temperature $T_1 \approx V_{(0)}^{1/4}$, and its further evolution proceeds in a standard way [10] [+1].

A most detailed discussion of this scenario is contained in a very interesting paper of Guth [12], where it is shown that the existence of a sufficiently long period of exponential expansion (inflation) in the early universe would provide a natural solution of the horizon and flatness problems in cosmology and of the primordial monopole problem in grand unified theories [13].

Unfortunately, however, this scenario in the form suggested in ref. [12] leads to some unacceptable consequences, recognized by Guth himself and by other authors who have studied this problem later, see e.g. refs. [8,14–18]. The phase transition from the symmetric vacuum state $\varphi = 0$ to the asymmetric state $\varphi = \varphi_0$ proceeds by creation and subsequent expansion of bubbles containing some nonvanishing fields φ. In ref. [12] it was implicitly assumed that inside these bubbles the scalar field φ rapidly grows to $\varphi = \varphi_0$, all energy of the bubbles becomes concentrated in their walls and thermalization occurs only after the collision of the walls. If this qualitative picture were correct, the exponential expansion would be finished at the temperature $T_{\rm c}$, at which the phase transition occurs. For the flatness problem to be solved the universe (the scale factor a) should grow at least 10^{28}

[+1] For an alternative scenario of the exponential expansion at the very early stages of the evolution of the universe, which may occur due to quantum gravity effects, see ref. [11].

0 031-9163/82/0000–0000/$ 02.75 © 1982 North-Holland

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

times during the exponential expansion period [12]. Since at this period the value of aT is constant, the critical temperature T_c should be 10^{28} times smaller than the temperature T_0, at which the exponential expansion starts. In the simplest SU(5) model [19] $T_0 \sim 10^{14}$ GeV, so that

$$T_c \lesssim 10^{-14}\ \text{GeV} \sim 0.1\ \text{K}. \quad (1)$$

No GUTs with such a fantastically small value of the critical temperature have been suggested.

There is also another problem with the above mentioned scenario. If the bubble wall collisions are necessary for the reheating of the universe, then after such a phase transition the universe becomes greatly inhomogeneous and anisotropic, which would contradict cosmological data [8,12,17,18].

In the present paper we would like to suggest an improved inflationary universe scenario, which is free of the above mentioned difficulties. With this purpose we shall consider the phase transitions in GUTs with the Coleman–Weinberg mechanism of symmetry breaking [20]. Phase transitions in such theories have been studied recently by many authors [14,15,21–23]. In our opinion, however, several important features of these phase transitions have escaped their attention. A detailed discussion of the phase transitions in GUTs with the Coleman–Weinberg mechanism of symmetry breaking will be contained in a subsequent publication. Here we shall only outline the main idea, which is essential for the understanding of the new inflationary universe scenario.

For definiteness let us consider the SU(5) grand unified theory [19], though most of what will be discussed here will not depend on the details of the model under consideration. The one-loop effective potential for the symmetry breaking SU(5) → SU(3) × SU(2) × U(1) in the Coleman–Weinberg version of this model at finite temperature T is [1–3, 14,15]

$$V(\varphi, T) = (18T^4/\pi^2) \times \int_0^\infty dx\, x^2 \ln\{1 - \exp[-(x^2 + 25g^2\varphi^2/8T^2)^{1/2}]\} + (5625/512\pi^2)g^4(\varphi^4 \ln(\varphi/\varphi_0) - \varphi^4/4 + \varphi_0^4/4), \quad (2)$$

where $\varphi_0 \sim 10^{14}$–10^{15} GeV, $g^2 \sim 1/3$ is the gauge coupling constant. At $T \gg \varphi_0$ the symmetry in this theory was restored, $\varphi = 0$ [1–3]. With a decrease of temperature the absolute minimum of $V(\varphi)$ appears at $\varphi \approx \varphi_0$. However at any $T \neq 0$ the point $\varphi = 0$ remains a local minimum of $V(\varphi, T)$, since near $\varphi = 0$

$$V(\varphi, T) = \tfrac{75}{16}g^2T^2\varphi^2 - (5625/512\pi^2)g^4\varphi^4 \ln(M_x/T) + (9/32\pi^2)M_x^4, \quad (3)$$

where $M_x^2 = \frac{25}{8}g^2\varphi_0^2$. The phase transition with symmetry breaking proceeds from a strongly supercooled state $\varphi = 0$ at temperature T_c, which is many orders of magnitude smaller than φ_0 [14,15,21–23]. The shape of the potential $V(\varphi, T)$ for $T \ll \varphi_0$ is shown in fig. 1. The phase transition begins with the formation of bubbles of the field φ, which is a tunneling process [6]. One may argue that for $T_c \ll \varphi_0$ this process does not depend on the properties of $V(\varphi, T_c)$ at $\varphi \sim \varphi_0$, and the maximal value of the field φ inside the bubble immediately after its formation should be of the order of φ_1, where

$$V(\varphi_1, T_c) = V(0, T_c), \quad (4)$$

$\varphi_1 \ll \varphi_0$, see fig. 1. Indeed, a detailed study of the bubble formation in this theory performed in refs. [24,25] by means of computer calculations shows that at the moment of the bubble formation the maximal value of the field φ inside the bubble equals approximately $3\varphi_1$. Therefore inside the bubble

$$\varphi \lesssim 3\varphi_1 = \frac{12\pi T_c}{5g}\left(\frac{2}{3\ln(M_x/T_c)}\right)^{1/2} \ll \varphi_0. \quad (5)$$

This means that the (negative) mass squared of the

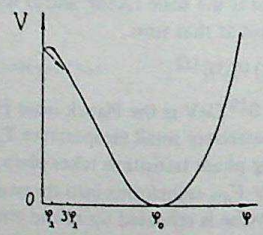


Fig. 1. Effective potential in the Coleman–Weinberg theory for $T \ll \varphi_0$. The arrow indicates the direction of the tunneling with bubble formation.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

field φ inside the bubble is

$$-m^2 = -\tfrac{2}{15}\,\mathrm{d}^2V/\mathrm{d}\varphi^2 \lesssim 75g^2T_c^2 \sim 25T_c^2. \quad (6)$$

After the bubble formation the field φ inside the bubble gradually grows up to its equilibrium value $\varphi(T_1) \sim \varphi_0$. At the first stages of this process the field φ grows approximately as e^{mt}. Therefore it approaches its equilibrium value $\varphi(T_1)$ only after some period of time $\tau \gtrsim m^{-1} \sim 0.2T_c^{-1}$. A more complete investigation shows that τ is several times greater than m^{-1}; here for simplicity we shall take as an estimate

$$\tau \sim T_c^{-1}. \quad (7)$$

It can also be easily shown that during most of this period the field φ inside the bubble remains much less than φ_0. Therefore during some time of the order of $\tau \sim T_c^{-1}$ the vacuum energy density $V(\varphi)$ remains almost equal to $V(0)$, and the part of the universe inside the bubble expands exponentially just as it expanded before the bubble creation. This simple observation has very important consequences for the theory of the phase transitions in the Coleman–Weinberg model.

Let us suppose that the phase transition in the Coleman–Weinberg SU(5) theory occurs at $T_c \sim 2 \times 10^6$ GeV, as it was claimed in ref. [15]. From eq. (3) it follows that with the parameters of the theory used in ref. [15] ($M_x \sim 6 \times 10^{14}$ GeV) the Hubble constant

$$H = [(8\pi/3M_P^2)V(0)]^{1/2}$$

is equal to 1.5×10^{10} GeV. Therefore during the exponential expansion period $\tau \sim T_c^{-1}$ the universe should grow $e^{H\tau}$ times, where

$$e^{H\tau} \sim e^{H/T_c} \sim e^{7500} \sim 10^{3260}. \quad (8)$$

A typical size of the bubble at the moment of its creation is $O(T_c^{-1}) \sim 10^{-20}$ cm [25]. After the period of the exponential expansion this bubble will have a size of

$$10^{-20} \cdot e^{H\tau}\ \mathrm{cm} \sim 10^{3240}\ \mathrm{cm},$$

which is much greater than the size of the observable part of the universe $l \sim 10^{28}$ cm. Therefore the whole observable part of the universe is contained *inside one bubble*, so we see no inhomogeneities caused by the wall collisions. After some time of the order of τ after the bubble creation all the vacuum energy density $V(0)$ transforms into thermal energy $\sim T_1^4$, where in our model $T_1 \approx 0.15M_x \sim 10^{14}$ GeV. However the thermalization occurs now not due to the wall collisions, but due to the interactions of particles created by the classical homogeneous field φ, convergently oscillating near its equilibrium value $\varphi(T_1) \approx \varphi_0$ with a frequency of about 10^{14} GeV.

One can easily verify that the size of the particle horizon at the time of the phase transition was much greater than the size of the bubble $\sim T_c^{-1}$, i.e. all points inside the bubble were causally connected. After the exponential expansion period this causally connected domain covers the whole observable part of the universe, which solves the horizon problem [12].

Now let us remember that particle creation in the very early universe in general cannot make it completely isotropic, but makes it quasi-isotropic [10], i.e. locally isotropic in small domains of space of the size of the same order as or greater than the Planck length $l_P \sim 10^{-33}$ cm $\sim M_P^{-1}$ at the Planck time $t_P \sim 10^{-43}$ s, when the temperature was $T_P \sim M_P \sim 10^{19}$ GeV. Since before the phase transition the quantity aT was constant inside each isotropic domain of the universe, at the moment of the phase transition the typical size of the isotropic domain exceeds the bubble size $\sim T_c^{-1}$. Therefore the space–time inside the bubble was isotropic and the exponential expansion extends this isotropy to the whole observable part of the universe. (Moreover, the remaining small anisotropy inside the bubble decreases rapidly during the exponential expansion period [18].) This may solve the long-standing problem of the space–time isotropy in our universe.

Density fluctuations inside the bubble immediately after its formation are negligibly small as compared with $V(0)$, i.e. the space inside the bubble is almost homogeneous. Then the exponential expansion extends this homogeneity to the whole observable part of the universe, which explains the large-scale homogeneity of the universe.

One may argue that it is not very good to obtain an absolutely homogeneous universe, since in that case it would be difficult to understand the origin of galaxies. However, as will be explained in a separate publication (see also refs. [7,26]), the necessary inhomogeneities may be generated after a subsequent

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

phase transition with a smaller degree of supercooling. Moreover, as is shown in ref. [27], the perturbations necessary for galaxy formation arise due to quantum gravity effects just after phase transitions of the type considered above in GUTs with the unification scale $\Lambda \sim 10^{17}-10^{18}$ GeV, which is not unrealistic [28].

From our results it follows that the size of the universe l_1 after the phase transition should exceed 10^{3240} cm, and the temperature T_1 is of the order of 10^{14} GeV. Therefore the total entropy of the universe should exceed $(l_1 T_1)^3 \sim 10^{10000}$, which explains why the total entropy of the universe exceeds 10^{85} and simultaneously solves the flatness problem [12,21].

It is known that the primordial monopoles in GUTs are created only in the points, in which bubbles with different types of Higgs field φ collide [13]. Therefore in our scenario no monopoles are created in the observable part of the universe, which solves the primordial monopole problem in GUTs [13]. For the same reason there will be no domain walls in the observable part of the universe in the theories with broken discrete symmetries [29], and in particular in the theories with spontaneously broken *CP* invariance. This helps to solve the problem of the baryon asymmetry of the universe [30]. Moreover, in the scenario under consideration there appears an additional source of baryon asymmetry. The standard mechanism is connected with the decay of the X, Y bosons and Higgs mesons, which appear in the course of the reheating of the universe during the phase transition. It is clear, however, that the baryon asymmetry generated by the decay of the Higgs mesons may be generated by the decay of the classical Higgs field vacuum $\varphi = 0$ as well. Note also that, whereas in a standard scenario the particles created by the decay of the X, Y and Higgs mesons are only a few percent of the total amount of particles [30], in our case all particles which appear after the phase transition are created by the oscillating classical Higgs field φ during the phase transition.

The new inflationary universe scenario discussed above is, of course, oversimplified. To get a complete scenario, one should analyse the phase transitions in the Coleman–Weinberg model more accurately taking into account the renormalization group equation [21] and the nonperturbative effects [15]. This analysis should be performed simultaneously with the investigation of the effects connected with the nonvanishing curvature and rapid expansion of the universe, which become important for $H > T_c$. One may ask e.g. whether it is possible for the universe to be in a state with temperature T_c smaller than the Hawking temperature $T_H = H/2\pi$, to which consequences the terms $\sim R\varphi^2$ in the effective potential may lead etc. [23]. An investigation of these problems is rather involved and will be contained in a separate publication. Our preliminary result is that there exists an *improved Coleman–Weinberg theory*, in which $d^2V/d\varphi^2 = 0$ at $\varphi = 0$ not in Minkowski space, but rather in de Sitter space with the curvature R determined by the vacuum energy density $V(\varphi)$ at the symmetric point $\varphi = 0$. In some versions of this theory the phase transition with symmetry breaking occurs due to nonperturbative effects [15]. The kinetics of this phase transition is somewhat more complicated than that described above, but the main feature of the new inflationary universe scenario remains intact: The field φ approaches its equilibrium value $\varphi(T_1) \sim \varphi_0$ during the period $\tau \gg H^{-1}$, which just leads to the desirable inflation of the universe discussed in the present paper.

I would like to express my deep gratitude to G.V. Chibisov, P.C.W. Davies, V.P. Frolov, L.P. Grishchuk, S.W. Hawking, R.E. Kallosh, D.A. Kirzhnits, V.F. Mukhanov, V.A. Rubakov, A.A. Starobinsky, A.V. Veryaskin and Ya.B. Zeldovich for many enlightening discussions.

References

[1] D.A. Kirzhnits, JETP Lett. 15 (1972) 529;
D.A. Kirzhnits and A.D. Linde, Phys. Lett. 42B (1972) 471;
S. Weinberg, Phys. Rev. D9 (1974) 3357;
L. Dolan and R. Jackiw, Phys. Rev. D9 (1974) 3320;
D.A. Kirzhnits and A.D. Linde, Zh. Eksp. Teor. Fiz. 67 (1974) 1263 [Sov. Phys. JETP 40 (1975) 628].
[2] D.A. Kirzhnits and A.D. Linde, Ann. Phys. (NY) 101 (1976) 195.
[3] A.D. Linde, Rep. Prog. Phys. 42 (1979) 389.
[4] A.D. Linde, in: Statistical mechanics of quarks and hadrons, ed. H. Satz (North-Holland, Amsterdam, 1981) p. 385; Phys. Lett. 99B (1981) 391.
[5] M. Daniel, Phys. Lett. 98B (1981) 371.
[6] A.D. Linde, Phys. Lett. 70B (1977) 306; 100B (1981) 37.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

[7] K. Sato, Mon. Not. R. Astron. Soc. 195 (1981) 467.
[8] A.H. Guth and E. Weinberg, Phys. Rev. D23 (1981) 876.
[9] R. Tolman, Relativity, thermodynamics and cosmology (Clarendon, Oxford, 1969).
[10] Ya.B. Zeldovich and I.D. Novikov, Structure and evolution of the universe (Nauka, Moscow, 1975).
[11] A.A. Starobinsky, Phys. Lett. 91B (1980) 100; Ya.B. Zeldovich, Pis'ma Astron. Zh. 7 (1981) 579.
[12] A.H. Guth, Phys. Rev. D23 (1981) 347.
[13] Ya.B. Zeldovich and M.Yu. Khlopov, Phys. Lett. 79B (1978) 239; J.P. Preskill, Phys. Rev. Lett. 43 (1979) 1365.
[14] G.P. Cook and K.T. Mahanthappa, Phys. Rev. D23 (1981) 1321.
[15] A. Billoire and K. Tamvakis, CERN preprint TH. 3019 (1981); K. Tamvakis and C.E. Vayonakis, CERN preprints TH. 3108, TH. 3128 (1981).
[16] Q. Shafi, CERN preprint TH. 3143 (1981).
[17] S.W. Hawking, I.G. Moss and J.M. Stewart, DAMTP preprint (1981).
[18] J.D. Barrow and M.S. Turner, Nature 292 (1981) 35.
[19] H. Georgi and S.L. Glashow, Phys. Rev. Lett. 32 (1974) 389.
[20] S. Coleman and E. Weinberg, Phys. Rev. D7 (1973) 1888.
[21] V.G. Lapchinsky, V.A. Rubakov and A.V. Veryaskin, IYal preprint (1981).
[22] M. Sher, Univ. of California preprint NSF-ITP-81-15 (1981).
[23] L.F. Abbott, Nucl. Phys. B185 (1981) 233; P. Hut and F.R. Klinkhamer, Phys. Lett. 104B (1981) 439.
[24] E. Brezin and G. Parisi, J. Stat. Phys. 19 (1978) 269.
[25] A.D. Linde, Decay of the false vacuum at finite temperature, Lebedev Phys. Inst. preprint (1981).
[26] Ya.B. Zeldovich, Mon. Not. R. Astron. Soc. 192 (1980) 663; A. Vilenkin, Phys. Rev. Lett. 46 (1981) 1169; Tufts Univ. preprint (1981).
[27] G.V. Chibisov and V.P. Mukhanov, Lebedev Phys. Inst. preprint No. 198 (1981); Mon. Not. R. Astron. Soc., to be published; D.A. Kompaneets, V.N. Lukash and I.D. Novikov, Space Research Inst. preprint No. 652 (1981).
[28] S. Dimopoulos, S. Raby and F. Wilczek, Stanford Univ. preprint NSF-ITP-81-31 (1981); S. Dimopoulos and H. Georgi, Harvard Univ. preprint HUTP-81/A022 (1981).
[29] Ya.B. Zeldovich, I.Yu. Kobzarev and L.B. Okun, Sov. Phys. JETP 40 (1975) 1.
[30] A.D. Sakharov, Pis'ma Zh. Eksp. Teor. Fiz. 5 (1967) 32; M. Yoshimura, Phys. Rev. Lett. 41 (1978) 281; S. Dimopoulos and L. Susskind, Phys. Rev. D18 (1978) 4500; J. Ellis, M.K. Gaillard and D.V. Nanopoulos, Phys. Lett. 80B (1979) 360; S. Weinberg, Phys. Rev. Lett. 42 (1979) 859.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

SUPERCOOLED PHASE TRANSITIONS IN THE VERY EARLY UNIVERSE

S.W. HAWKING and I.G. MOSS
Department of Applied Mathematics and Theoretical Physics, Cambridge, UK

Received 12 January 1982

The universe might have had a prolonged exponentially expanding phase caused by its being stuck in a metastable state of the grand unified phase transition. The only way that it could exit from this exponential expansion without introducing too much inhomogeneity or spatial curvature would be through a homogeneous "bubble" solution in which quantum tunnelling occured everywhere at the same time. This would produce more baryons than the conventional scenarios.

There has been considerable interest recently in the possibility that the universe might have undergone an exponentially expanding "inflationary" stage caused by the vacuum energy of a supercooled metastable state of a first-order phase transition associated with the breakdown of the grand unification symmetry [1–4]. Such an exponential expansion might provide the answer to a number of problems in cosmology such as

(1) Why are there so few magnetic monopoles [1,5].

(2) Why is the universe so spatially homogeneous and isotropic?

(3) Why is the universe so nearly flat, i.e. why is it expanding at almost exactly the critical rate to avoid recollapse?

The main difficulty with this inflationary scenario has been to find some way of getting out of the exponentially expanding stage without creating large-scale inhomogeneities that would be incompatible with observation. In his original paper Guth [2] suggested that small bubbles of the broken symmetry phase would form by vacuum tunelling [6] and then expand by a large factor before colliding with other bubbles. However, as he himself admitted, and others have confirmed [7], one would end up with a very inhomogeneous universe dominated by a few very large bubbles.

Most of the papers that have discussed the exit from the inflationary stage have treated it essentially as a phase transition problem in flat spacetime and have neglected the curvature and finite horizon size of the universe. The aim of this letter is to show that these effects play a very important role and can lead to a phase transition which occurs simultaneously at all points of space thus not creating any inhomogeneity or magnetic monopoles. If the adjustable parameters in the effective potential for the Higgs fields lie in a certain range, the spatial flatness of the universe would also be accounted for.

The effective potential of the Higgs fields has the general form

$$V(\phi) = \tfrac{1}{2}(m^2 + \xi R + cT^2)\phi^2 + \tfrac{1}{4}\alpha^2\phi^4(\log \phi^2/\phi_0^2 - \tfrac{1}{2} + m^2/\alpha^2\phi_0^2) + \tfrac{1}{8}\alpha^2\phi_0^4 - \tfrac{1}{4}m^2\phi_0^2 , \quad (1)$$

where ϕ is some measure of the magnitude of the Higgs field and ϕ_0 is its expectation value at zero temperature in flat space with $V(\phi_0) = 0$. The quanties m^2 and ξ are renormalised parameters which can be given any value. $\xi = \frac{1}{6}$ is a natural choice because it corresponds to conformal invariance when $m = 0$ and it is not renormalized at one loop though it is at two loops. The quantity c is a numerical factor of order 1 which depends on the group and the

0 031-9163/82/0000–0000/$ 02.75 © 1982 North-Holland

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

representation. Neglecting small corrections arising from the ξR term, the energy density of the universe will be

$$\rho = (\pi^2/30)g(T)T^4 + V(\phi,T)\,, \tag{2}$$

where $g(T)$ is the effective number of spin states of relativistic particles. The rate of expansion will be given by

$$H^2 = (8\pi/3m_\rho^2)\rho - k/S^2 \tag{3}$$

where $H = \dot{S}/S$ and S is the scale factor of the universe.

At very high temperatures the universe will be in the symmetric phase $\phi = 0$. If $m^2 + \xi R \geqslant 0$, the symmetric phase will remain a local minimum of the potential as the temperature falls although it will cease to be a global minimum at some critical temperature T_c of the order of $\alpha\phi_0$. However, in realistic models the probability to tunnel to the broken symmetry phase with $\phi \approx \phi_0$ is very low [8]. The universe will therefore continue to expand in the symmetric phase until the thermal energy, the first term in eq. (2), becomes small compared to the vacuum energy $V(0, T)$. This will lead to an exponential expansion and the universe will rapidly approach a de Sitter state which depends only on the value of $V(0)$ through the relation

$$H = [8\pi V(0)/3m_p^2]^{1/2} \tag{4}$$

but which is otherwise independent of the initial conditions and the initial value of k. This is very similar to the way that a gravitational collapse rapidly approaches a stationary black-hole state which depends only on the mass and angular momentum but which is otherwise independent of the nature of the collapsing body. Following the black-hole analogy, we shall therefore assume a cosmological "no hair" theorem and treat the universe as a de Sitter model.

In the gravitational collapse case, the black hole has an effective temperature [9] $T_B = m_p^2(8\pi M)^{-1}$. Similarly, de Sitter space has an effective temperature [10] $T_s = H/2\pi$. In the case of grand unified phase transitions this will be of the order of 10^{11} GeV. In a de Sitter universe it is impossible to define a temperature smaller than T_s. It is therefore incorrect to use in the effective potential (1) the redshifted temperature of the universe when this falls below T_s. In fact one can regard the ξR term as representing the contribution of T_s to the effective potential.

In accordance with our viewpoint that de Sitter space has no hair, we shall calculate the probability of tunnelling from the local minimum at $\phi = 0$ to the global minimum by looking for solutions of the classical coupled scalar and gravitational field equations that are near the euclidean version of de Sitter space, i.e. a four-sphere of radius H^{-1}.

If $m^2 > 2H^2(1-6\xi)$, there are inhomogeneous "bubble" solutions as there are in the decay of the false vacuum in flat spacetime [6]. Gravity and the curvature of the universe do not have such effect except to make the bubble slightly smaller [11]. However, as mentioned earlier, this case would lead to unacceptably large inhomogeneities in the universe today.

If $2H^2(1 - 6\xi) \geqslant m^2 \geqslant -12H^2\xi$, the potential (1) would lead to inhomogenous bubble solutions in flat spacetime whose radius was greater than that of the de Sitter space, H^{-1}. In this case the only euclidean solution of the coupled equations, apart from $\phi = 0$, is the homogeneous solution $\phi = \phi_1$ where $V(\phi_1)$ is the local maximum of V on a four-sphere of radius H_1^{-1} where $H_1^2 = 8\pi V(\phi_1)/3m_p^2$. The tunnelling probability per unit four-volume is of the order of $(m^2 + \xi R)^2 e^{-B}$ where

$$B = \tfrac{1}{8}m_p^4\,[1/V(0) - 1/V(\phi_1)] \tag{5}$$

is the difference between the combined gravitational and scalar field actions of the $\phi = \phi_1$, and the $\phi = 0$ solutions. Even though B may be very large and the tunnelling probability very small, this does not matter because the universe will continue in the essentially stationary de Sitter state until it makes a quantum transition everywhere to the $\phi = \phi_1$ solution.

The $\phi = \phi_1$ solution is unstable as is reflected by the fact that it has a negative perturbative mode $\phi = \phi_1 + \epsilon$ where ϵ is a constant on the four-sphere. One would therefore expect the solution in lorentzian spacetime to evolve according to the classical field equations with ϕ running downhill from the maximum at $\phi = \phi_1$ to the global minimum at $\phi = \phi_0$. The solution will presumably be spatially homogeneous but evolving in time so it can be represented by a Friedmann–Robertson–Walker metric with ϕ a functionof time only. One can introduce Robertson–Walker coordinates with $K = +1, 0$ or -1 into de Sitter space though

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

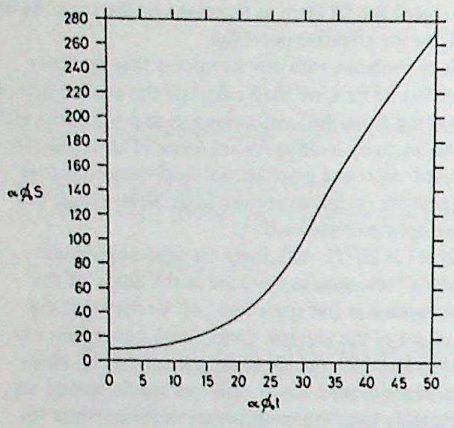


Fig. 1. The scale factor of the universe starting from a $k = +1$ time symmetric state with $\phi = \phi_1 + \epsilon$.

the coordinates cover the whole space only in the first case. It would seem reasonable to start the classical evolution of the solution on a spacelike surface whose curvature (positive or negative) is of the order of the radius H_1^{-1} of the de Sitter space. The most natural choice would seem to be the surface of time symmetry in the $K = +1$ coordinates but the conclusions in the $K = -1$ case are very similar.

While the scalar field remains near $\phi = \phi_1$, the expansion of the universe will continue to be nearly exponential or, in the $K = +1$ case, $S = H_1^{-1} \cosh H_1 t$ where S is the scale factor. The timescale for ϕ to run downhill from the maximum at $\phi = \phi_1$ is

$$3H_1/(m^2 + 12\xi H_1^2)\,. \tag{6}$$

When ϕ approaches the global minimum ϕ_0 it will start to perform damped harmonic osciallations around $\phi = \phi_0$ and the expansion of the universe will change to that of a model filled with non relativistic matter. Figs. 1 and 2 show the numerical solutions for S and ϕ in the case $K = +1$, $m^2 + 12\xi H_1^2 = 24 H_1^2$.

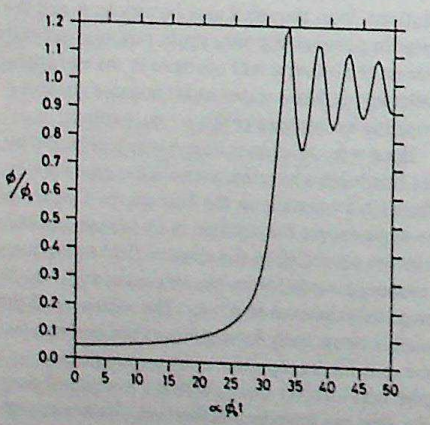


Fig. 2. The value of the scalar field in the same universe.

One can think of the oscillations of ϕ about ϕ_0 as corresponding to a very large number of Higgs particles in a coherent state. Presumably they will decay into light gauge particles and fermions which will have an approximately thermal distribution at a temperature T_2 given by $(\pi^2/30)g(T_2)T_2^4 \approx V(\phi_1)$. The decay of the Higgs particles would probably give a much larger baryon asymmetry than in the normal scenario.

The radius of the curvature of the space sections of the universe at the time of thermalization will be of the order $H_1^{-1} \cosh H_1 t_2$ where t_2 is the time from the $\phi = \phi_1$ state to thermalization. If one assumes that this is a few times the timescale (6) for ϕ to run downhill, one finds that the universe would have been sufficiently spatially flat to expend to its present size with a nearly critical density provided that $m^2 + 12\xi H_1^2 < \frac{1}{12} H_1^2$.

This seems the only situation in which one can have a significant exponential expansion at the GUT era without introducing either too much inhomogeneity or too much spatial curvature.

Finally, if $m^2 + 12\xi H_1^2 < 0$, the symmetric state $\phi = 0$ will cease to be a local minimum of the potential before the temperature drops to T_s so there will not be a prolonged exponential expansion phase.

We are very grateful to A.D. Linde for discussions which gave rise to this paper.

References

[1] A.H. Guth and S.H. Tye, Phys. Rev. Lett. 44 (1980) 631, 963.
[2] A.H. Guth, Phys. Rev. D23 (1981) 347.
[3] A.H. Guth and E.J. Weinberg, Phys. Rev. D23 (1981) 876.
[4] A.D. Linde, Lebedev Institute preprint 229 (1981).
[5] Ya.B. Zeldovich and M.Yu Khlopov, Phys. Lett. 79B (1978) 239.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

J. Preskill, Phys. Rev. Lett. 43 (1979) 1365.
[6] S. Coleman, Phys. Rev. D15 (1977) 2929;
C. Callan and S. Coleman, Phys. Rev. D16 (1977) 1762.
[7] S.W. Hawking, I.G. Moss and J.M. Stewart, Bubble collisions in the very early universe, Cambridge preprint (1981).
[8] A.D. Linde, Phys. Lett. 99B (1981) 391.
G.P. Cook and K.T. Mahanthappa, Phys. Rev. D23 (1981) 1321;
A. Billoire and K. Tamvakis, CERN preprint TH.3019 (1981)
[9] S.W. Hawking, Commun. Math. Phys. 43 (1975) 199.
[10] G.W. Gibbons and S.W. Hawking, Phys. Rev. D15 (1977) 2738.
[11] S. Coleman and F. De Luccia, Phys. Rev. D21 (1980) 3305.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking

Andreas Albrecht and Paul J. Steinhardt
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104
(Received 25 January 1982)

The treatment of first-order phase transitions for standard grand unified theories is shown to break down for models with radiatively induced spontaneous symmetry breaking. It is argued that proper analysis of these transitions which would take place in the early history of the universe can lead to an explanation of the cosmological homogeneity, flatness, and monopole puzzles.

PACS numbers: 98.80.Bp, 11.15.Ex, 12.10.En

Hot big-bang cosmology depends upon special conditions for the early universe to explain the high degree of homogeneity (the "homogeneity puzzle")[1] and the nearly critical mass density (the "flatness puzzle")[2] found in the universe today. In addition, it has been shown that in typical grand unified theories (GUT's) phase transitions should occur in the early history of the universe which lead to many more magnetic monopoles being produced and surviving to the present epoch than are consistent with experiment (the "monopole puzzle").[3]

In this paper we will argue that first-order phase transitions in a special class of GUT's —models in which the GUT symmetry is broken by radiatively induced corrections to the tree approximation to the effective potential—can lead to a solution to these and other cosmological puzzles. (Models with radiatively induced symmetry breaking, a mechanism discovered by Coleman and Weinberg,[4] will be referred to as CW models.) In particular, we will present results for the standard GUT with a finite-temperature effective (scalar) potential:

$$V_T(\varphi) = (2A - B)\sigma^2\varphi^2 - A\varphi^4 + B\varphi^4 \ln(\varphi^2/\sigma^2) + 18(T^4/\pi^2)\int_0^\infty dx\, x^2 \ln\{1 - \exp[-(x^2 + 25g^2\varphi^2/8T^2)^{1/2}]\}, \quad (1)$$

where the adjoint Higgs field, Φ, has been reexpressed as $\varphi(1,1,1,-\frac{3}{2},-\frac{3}{2})$ (the fundamental Higgs field will be irrelevant for this discussion); g is the gauge coupling constant; σ is chosen to be 4.5×10^{14} GeV; $B = 5625g^4/1024\pi^2$; and A is a free parameter. Equation (1) includes the one-loop quantum and thermal corrections to the effective potential. For a CW model, the coefficient of the quadratic term, $2A - B$, is set equal to zero and the Higgs mass is $m_{CW} = 2.7\times10^{14}$ GeV. We will also present results for non-CW models in which $2A - B$ is small and therefore the Higgs mass, m_H, is such that $\Delta_H \equiv (m_H^2 - m_{CW}^2)/m_{CW}^2$ is small.

As for more general GUT models, the process of the first-order phase transition from the SU(5) *symmetric* phase to the SU(3)⊗SU(2)⊗U(1) *symmetry-breaking* phase for the CW model can be understood by studying the shape of the effective potential as a function of the scalar field for various values of the temperature, as shown in Fig. 1. For temperatures above the critical temperature (T_{GUT}) for the transition, the symmetric phase ($\varphi = 0$) is the global stable minimum of the effective potential. At $T = T_{GUT}$, the symmetric phase and the symmetry-breaking phase have equal energy densities. As the temperature drops below T_{GUT}, the symmetric phase becomes *metastable*—it has a higher-energy density than *stable* symmetry-breaking phase but a potential barrier prevents it from becoming unstable.

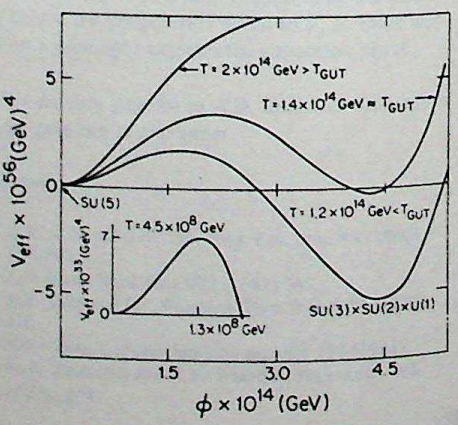


FIG. 1. Effective potential vs φ for various values of T.

© 1982 The American Physical Society

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

The decay of a metastable phase to a stable phase has been compared to a classical nucleation process. At $T < T_{GUT}$ there is a rate per unit time per unit volume, $\Gamma(t)$, for producing finite-sized fluctuations containing stable phase—bubbles—within the metastable system. Once produced, the bubbles grow, coaelesce and convert the system to the stable phase. For cases where the barrier is sufficiently large, Coleman[5] and Linde[6] have found methods for computing $\Gamma(t)$ using a steepest descent (SD) approximation and found it to be of the form $S\exp[-F_f(T)/kT]$. S has the dimensions of (length)$^{-4}$ and k is Boltzmann's constant. $F_f(T)$ is the free energy associated with the bubble computed by SD approximation to be the dominant path across the potential barrier. For this SD bubble fluctuation, the value of $\langle\varphi\rangle$ varies from φ_f (on the stable-phase side of the barrier) in the center of the bubble to $\varphi = 0$ far from the bubble center. As T decreases, $\Gamma(t)$ increases and φ_f/σ decreases, where $\varphi = \sigma$ corresponds to the symmetry-breaking minimum.

The maintenance of a system in a metastable phase during a first-order phase transition as $T \lesssim T_{GUT}$ continues to decrease is known as supercooling. For phase transitions in the early universe, Guth and Tye,[7] taking into account the expansion of the universe, found the expression for the fractional volume $[p(T)]$ of the universe which at temperature T has decayed to the stable phase during supercooling. When $p(T)$, which depends upon $\Gamma(t)$, is of order unity, the decay is said to be *terminated*.

Guth[8] recently suggested that supercooling of first-order phase transitions of typical GUT models can lead to a solution of the cosmological puzzles. His idea was that the energy density of the universe, ρ, during supercooling is dominated (once $T \lesssim T_{GUT}/10$) by the $\rho_0 \approx T_{GUT}^4$, the difference in energy density between the metastable and stable phases. Then, ρ_0 can act as a cosmological constant in Einstein's equation for standard cosmology described by a Robertson-Walker metric:

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi}{3M_P^2}\rho \approx \frac{8\pi}{3M_P^2}T_{GUT}^4 \equiv t_{exp}^{-2}, \qquad (2)$$

where M_P is the Planck mass. The result is exponential growth of the scale factor, $R(t) \sim R_0 \times \exp(t/t_{exp})$. If the growth is continued long enough, each nearly homogeneous region of the universe experiences an expansion which, Guth showed, could explain the cosmological homogeneity, flatness, and monopole puzzles. The problem with Guth's scenario is that when this high degree of expansion can be arranged in typical GUT's (by adjusting free parameters) the rate of expansion of the universe dominates the rate of production and growth of bubbles; the bubbles never coalesce to complete the transition. Since our own universe exhibits the symmetry breaking of the stable phase, it would have to lie within a single, rare bubble, in which case it is difficult to understand how the high entropy found in our universe could have been generated.[9]

We claim that CW models and near-CW models in which $\Delta_H < 7\times 10^{-6}$ possess special properties that result in completion of the transition to the symmetry-breaking phase along with tremendous expansion. Initially, the analysis of the supercooling for $T < T_{GUT}$ proceeds as in more general GUT models. However, as has been pointed out previously,[9] two important features must be taken into account. Firstly, the GUT fine-structure constant, α ($= \frac{1}{45}$ at $T = T_{GUT}$), increases as a function of temperature[10] until at $T_\alpha \approx 10^6$ GeV, it is of order unity and the one-loop approximation to $V_T(\varphi)$ is no longer valid. Secondly, the prefactor in the expression for $\Gamma(t)$, S, is given by T^4 for CW models since, near $\varphi = 0$, the only parameter with the dimensions of length that affects the barrier and, thus, the decay, is T^{-1}.[9]

When these features are taken into account in the standard SD analysis, they combine to yield the peculiar behavior for $F_f(T)/kT$ and $p(T)$ that is illustrated in Fig. 2. The curves for $F_f(T)/kT$ have been terminated at temperatures, T_{term},

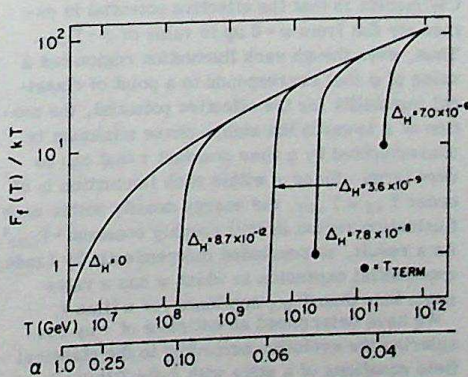


FIG. 2. $F_f(T)/kT$ vs T. Curves terminate for $p(T) \approx 1$. $\Delta_H = (M_H^2 - M_{CW}^2)/M_{CW}^2$.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

for which the fractional volume of stable phase, $p(T)$, is of order unity. For $\Delta_H > 7.0\times10^{-6}$, $F_f(T_{term})/kT_{term}$ is large, but there is insufficient expansion to solve the cosmological puzzles.

For smaller values of Δ_H, including CW models where $\Delta_H = 0$, $F_f(T_{term})/kT_{term}$ is of order unity or less, values which are too small to trust the SD approximation. The value of the temperature for which the SD approximation fails, T_{SD}, is a function of Δ_H and is roughly 10^8 GeV for $\Delta_H = 0$. T_{SD} is larger than T_α in all cases though, so that one can discuss the breakdown of the SD approximation while still considering only the one-loop approximation to the effective potential, Eq. (1).

For the CW and near-CW models, the fact that T reaches T_{SD} before the completion of the transition means that many other types of fluctuations (i.e., paths across the barrier) besides the SD bubble become important. The effect of the barrier becomes negligible. The system (the universe) can be thought of as balancing at a point of unstable (not metastable) equilibrium near $\varphi = 0$. Thermal fluctuations drive different regions of the universe away from the SU(5) symmetric phase but towards different symmetry-breaking minima. Since T is the only dimensional parameter relevant for the fluctuations, the average size of a fluctuation region should be of order T^{-1} and the (roughly constant) value of $\langle\varphi\rangle$ within a fluctuation region is of order T.

Even though $\langle\varphi\rangle \sim T$ corresponds to a point of the effective potential on the stable-phase side of the barrier, slightly to the right of the barrier in Fig. 1, a crucial feature of CW models and near-CW models is that the effective potential is extremely flat from $\varphi = 0$ up to value of $\varphi \sim T_{GUT}$. Thus, even though each fluctuation region has a value of φ that corresponds to a point of classical instability for the effective potential, the motion of φ towards the stable-phase minimum is characterized by a time constant τ that can be very large. Since φ within each fluctuation is of order $T_{SD} \ll T_{GUT}$, the energy density within each fluctuation region is still roughly constant $\sim T_{GUT}^4$. As a result, as concluded independently by Linde,[11] exponential expansion in which φ has a value much less than T_{GUT} continues for a time τ.

We have determined an estimate of τ by considering the evolution according to the classical field equations of a state with $\langle\varphi\rangle = T$ throughout space (presumably a similar method to what was used in Ref. 11) for a range of temperatures for fluctuation production. We found τ for CW models to be large compared to t_{exp}. This means that each fluctuation region undergoes many e foldings in spatial expansion before $\langle\varphi\rangle$ changes appreciably. Multiplying the scale factor after time τ by the average size of the initial fluctuation region (the size of an SD bubble was used for $T > T_{SD}$) we obtain the size of the fluctuation region, R_U, at time τ, after which $\langle\varphi\rangle \sim \sigma$ and the exponential expansion ceases. In Table I are shown the results for this computation for a range of temperatures. Column 2 shows R_U(flat) computed with use of the ordinary Klein-Gordon equation. Column 3 shows R_U(exp) derived by using the same equation with an extra drag term $(3\dot{R}\dot{\varphi}/R)$ included to account for the time dependence of the scale factor.[12] Column 4 shows R_U for $\Delta_H = 3.6\times10^{-11}$, where the time dependence of the scale factor has been included. For this value of Δ_H, the barrier disappears at $T = 3.7\times10^8$ GeV and the maximum value that R_U can achieve is $10^{-0.2}$ cm. Since the observed universe, according to the standard model, had a radius of ~1 cm for $T = 10^{14}$ GeV, the choice of Δ_H must be tuned to a value less than 3.6×10^{-11} in order for the observed universe to fit inside a single fluctuation region. Similarly, we have shown that if our scalar field couples to the curvature through a term $bC\varphi^2$ (C = curvature), $|b|$ must be $\lesssim 10^{-2}$. Please note, we have treated this calculation as if it were in flat space; curvature effects will be discussed in future publications.[12]

The result is that the size of a fluctuation region once $T < T_{SD}$ and the SD approximation breaks down is much greater than the size of our present observed universe (10^{28} cm). If one considers the "observed universe" as lying within such a fluctuation region of the "total universe" the special conditions of hot big-bang cosmology can be satisfied. Because the observed universe would be only a small portion of the total universe result-

TABLE I. Fluctuation radius vs T.

T (GeV)	R_U (flat) (cm)	R_U (exp) (cm)	R_U ($\Delta_H = 3.6\times10^{-11}$) (cm)
4.5×10^6	10^{81}	$\gg 10^{470}$	...
4.5×10^7	$10^{-4.6}$	10^{470}	...
1×10^8	10^{-13}	10^{98}	...
3×10^8	10^{-19}	$10^{-4.4}$	...
4.5×10^8	10^{-20}	10^{-13}	$10^{-7.2}$
4.5×10^9	10^{-23}	10^{-23}	10^{-23}

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ing from the extreme expansion of a small homogeneous region, the homogeneity puzzle is solved. The exponentially large value of scale factor accounts for the flatness puzzle.[8] Because $\langle\varphi\rangle \approx$ const over a fluctuation region, there is no reason to expect to find any monopoles (beyond a small thermally produced number) in the observed universe. Even though there is a discrete symmetry ($\varphi \to -\varphi$) in the theory, the distance between domain walls (separating regions with $\langle\varphi\rangle$ of opposite sign) produced in the transition should be greater than 10^{28} cm, and hence, unobservable. The potential energy stored in the scalar field is eventually converted to thermal energy,[13] thus producing a sizable entropy density inside each region. Thus, it appears that all the fundamental cosmological puzzles are solved.

A more complete discussion of these results[12] and an analysis of how a fluctuation evolves to thermal equilibrium and produces baryon asymmetry[13] will appear in forthcoming publications.

The authors thank A. Guth and E. Witten for useful discussions and G. Segre for his advice and support during the preparation of the manuscript. This work has been supported in part by the U.S. Department of Energy under Contract No. EY-76-C-02-3071.

[1]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).

[2]R. H. Dicke and P. J. E. Peebles, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge Univ. Press, Cambridge, 1979).

[3]J. Preskill, Phys. Rev. Lett. 43, 1365 (1979).

[4]S. Coleman and E. Weinberg, Phys. Rev. D 7, 788 (1973).

[5]S. Coleman, Phys. Rev. D 15, 2929 (1977).

[6]A. Linde, Phys. Lett. 70B, 306 (1977).

[7]A. Guth and S. H. H. Tye, Phys. Rev. Lett. 44, 631 963(E) (1980).

[8]A. Guth, Phys. Rev. D 23, 347 (1981).

[9]P. Steinhardt, in Proceedings of the 1981 Banff Summer Institute (to be published).

[10]M. Sher, Phys. Rev. D 24, 1699 (1981).

[11]A. Linde, Lebedev Institute Report No. 229, 1981 (to be published). Differences between our results and those of Linde will be discussed in a later paper. (A. Albrecht and P. Steinhardt, to be published).

[12]Albrecht and Steinhardt, Ref. 11.

[13]A. Albrecht, P. Steinhardt, M. Turner, and F. Wilczek, to be published.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

THE BOUNDARY CONDITIONS OF THE UNIVERSE

S.W. HAWKING
University of Cambridge, Cambridge

This paper considers the questions of what are the boundary conditions of the universe and where should they be imposed. It is difficult to define boundary conditions at the initial singularity and, even if one could, they would be insufficient to determine the evolution of the universe. In order to overcome this problem it is suggested that one should adopt the Euclidean approach and evaluate the path integral for quantum gravity over positive definite metrics. If one took these metrics to be compact, one would avoid the need to specify any boundary conditions for the universe. This approach might explain why the apparent cosmological constant is zero, why the universe is spatially flat, and why it was in thermal equilibrium at early times.

The aim of physics is to provide a mathematical model of the universe which will agree with all observations that have been made so far and which will predict the results of further observations. Our present models would consist of two parts:

i) a set of differential equations that govern the variables in the theory. These are normally derived from an action principle;

ii) boundary conditions for the differential equations, or for the fields that are considered in the action principle.

One could conceive of models which did not have this division into field equations and boundary conditions but the separation of the two has been very valuable in enabling us to make progress by considering only a

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

local region of the universe instead of having to try to devise a model which would account for the whole universe at one go. We have made a lot of progress on the first part of the problem in recent years and it now seems possible that we might find a fully unified field theory within the not-too-distant future. However, we shall not have a complete model of the universe until we can say more about the boundary conditions than that they must be whatever would produce what we observe.

In Newtonian theory, the gravitational potential obeys an elliptical equation in three-dimensional space at a given time. However, since the discovery of special relativity, we have believed that the fundamental fields all obey hyperbolic equations with the same characteristics, the null cone. This means that the boundary values can be taken to be data on a space-like surface. In the General Theory of Relativity the metric that defines the null cone is itself a dynamical field. Thus the causal structure of space-time may be very different from that of flat space. There may be closed or almost closed time-like curves such as shown in Figure 1. In this case the data on a space-like surface would have to be such that it reproduced itself exactly after a certain interval. This would be a stringent requirement which might have only one or a small number of solutions. I shall return to this kind of idea as a way of getting rid of boundary conditions.

Even if there are no closed time-like curves the universe might not

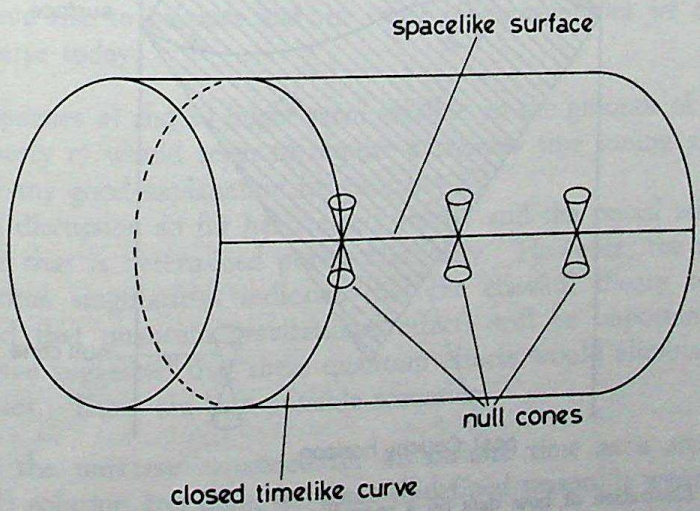


FIG. 1. An illustration of closed or almost closed time-like curves.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

be globally hyperbolic, i.e. there might not be any space-like surface such that data on that surface would determine the universe everywhere. A simple example is provided by anti-de Sitter space which is conformal to a time-like strip of Minkowski space. The conformal factor goes to infinity on the edges of the strip which are at infinity. As illustrated in Figure 2, data on a space-like surface determines the solution of the field equations only in a limited region, the Cauchy development, which is bounded by a null surface called the Cauchy horizon. Points beyond the Cauchy horizon have time-like curves through them which can go off to spatial infinity with-

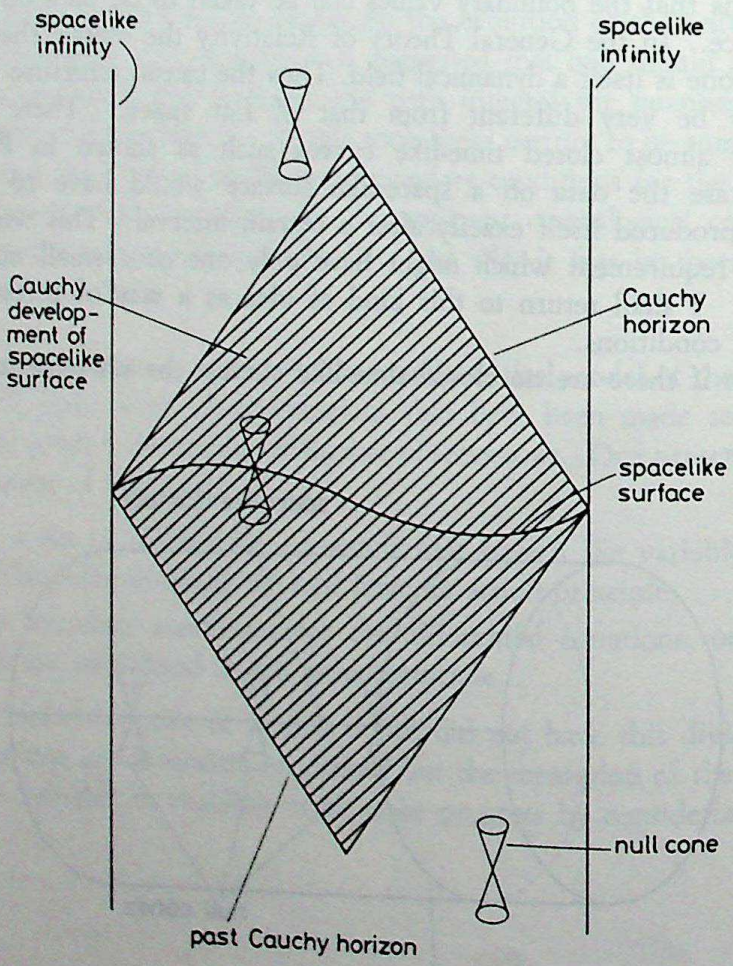


FIG. 2. An illustration of how data on a space-like surface determines the solution of field equations only within the Cauchy development.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

out intersecting the space-like surface. They can thus be influenced by data coming in from infinity.

Although there are a number of exact solutions such as anti-de Sitter space, the Reissner-Nordstrom and Kerr solutions which have Cauchy horizons, it seems likely that they are all unstable in the sense that a small perturbation of the initial data on the space-like surface will convert the Cauchy horizon into an unpassable singularity. This leads to the very plausible (though unproved) strong version of the Cosmic Censorship Hypothesis: generic initial data on a complete space-like surface produces a Cauchy development which is inextendible, i.e. it cannot be embedded in a larger non-singular spacetime. In other words the Cauchy development is either infinite in extent or it is bounded by singularities. In the case of a solution of the Einstein equations which was a model of our universe, we know from observation that there would be enough matter to cause it to be bounded in the past by a singularity, the Big Bang. It would seem natural to take this initial singularity as a place at which the boundary conditions of the universe were imposed. We would then have to explain why the universe started off:

a) spatially homogeneous and isotropic;

b) in thermal equilibrium;

c) spatially flat, i.e., with the kinetic energy of expansion exactly balanced by the negative gravitational potential energy;

d) notwithstanding *a*), *b*) and *c*), with small initial perturbations which gave rise to galaxies and the other inhomogeneities we observe in the universe today.

Properties *a*) and *b*) might seem sensible on the grounds of simplicity but property *c*) would seem to require extremely fine tuning and we do not have any good explanation for property *d*).

The discussion so far has treated gravity and the causal structure of spacetime that is determined purely classically. However, the prediction of spacetime singularities indicates that the classical theory will break down and that quantum gravitational effects will be important. Some people have suggested that these quantum effects would eliminate all the singularities. There are two possible scenarios:

i) the universe expanded for an infinite time as a steady state (de Sitter) solution and then, for some ill-defined reason, it went unstable and converted itself into the present day observed form;

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ii) there was a previous contracting phase followed by a non-singular "bounce".

I do not believe that quantum effects can remove all the singularities for the following reason. As shown in Figure 3 consider a star of a few solar masses containing about 10^{60} baryons which undergoes gravitational collapse to produce a black hole. Because of quantum effects the black hole will create and emit particles like a body with a temperature of about 10^{-7} K. At this low temperature the emission would consist almost entirely of zero rest-mass particles, such as neutrinos, photons and gravitons. These would carry away energy so that the mass of the black hole would slowly decrease. After a very long time the temperature of the black hole would become high enough for baryons and anti-baryons to be emitted in significant numbers. However by then the mass of the black hole would have fallen to about 10^{-20} of the original mass. Thus there would not be sufficient energy left to emit the original number of baryons. The semi-classical approximation used to derive the black hole emission will break down when the black hole gets down to near the Planck mass. We do not know how to treat this stage properly. There seem to be three possibilities:

i) the black hole turns into a naked singularity with a negative mass;

ii) the black hole stops emitting at about the Planck mass and remains as a stable object thereafter;

iii) the black hole disappears completely.

Possibilities i) and ii) would imply that quantum gravity was inherently asymmetric in time. One would expect the formation rate of black holes of mass M shortly after the Big Bang to be of order $M^4 \exp(-4\pi M^2/M_p^2)$ where M_p is the Planck mass. Thus one would expect large numbers of black holes of a few Planck masses. If these all left naked singularities or Planck mass black holes, they would completely dominate the mass density of the universe at the present time and would produce an enormous negative or positive deceleration parameter respectively. I therefore believe that black holes evaporate completely. If this is the case and if most of the original baryons do not reappear, there must be a naked singularity when the black hole finally disappears. This naked singularity will mean that spacetime is not globally hyperbolic. One would have expected this anyway because of quantum fluctuations of the light cone.

If quantum effects caused the universe not to be globally hyperbolic,

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

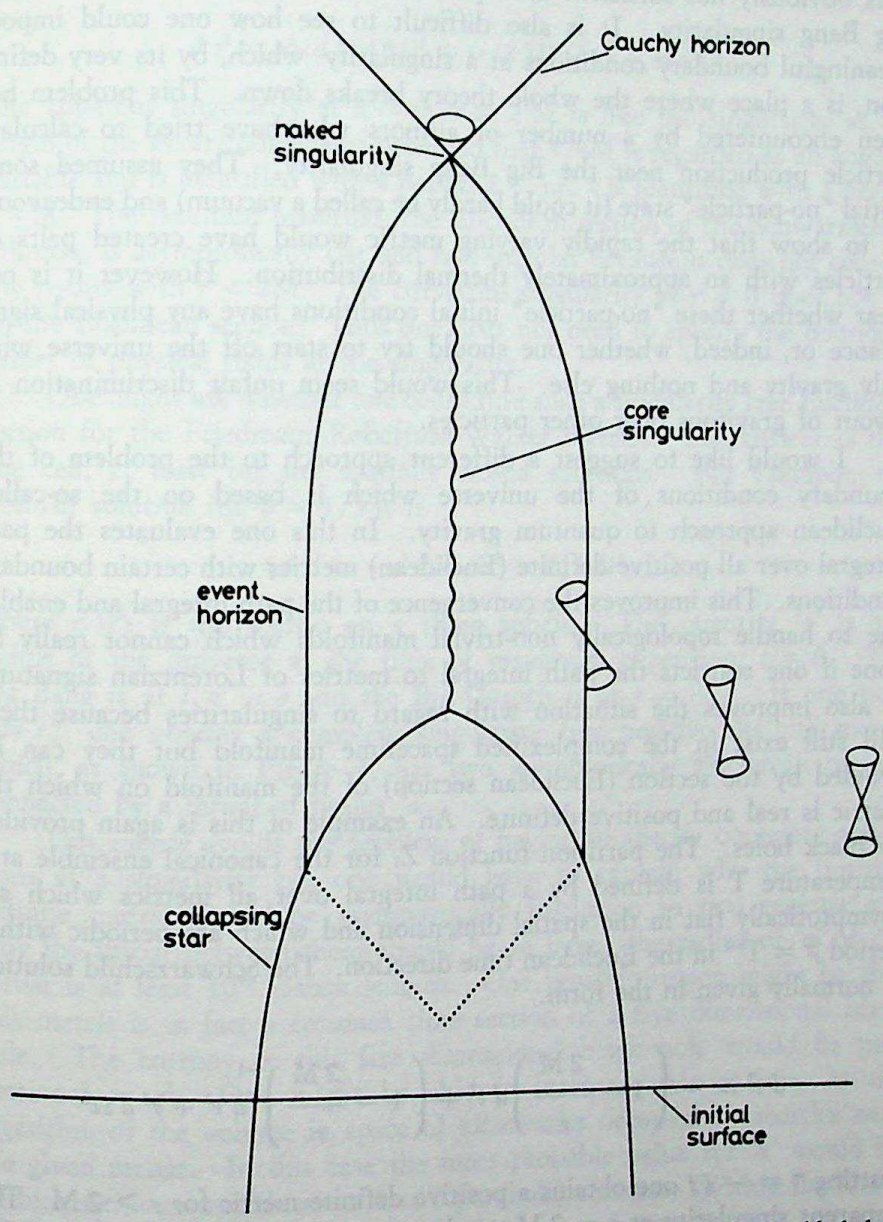


FIG. 3. A star of a few solar masses undergoes gravitational collapse to form a black hole which becomes a naked singularity.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

it is obviously not sufficient to impose boundary conditions at the initial Big Bang singularity. It is also difficult to see how one could impose meaningful boundary conditions at a singularity which, by its very definition, is a place where the whole theory breaks down. This problem has been encountered by a number of authors who have tried to calculate particle production near the Big Bang singularity. They assumed some initial "no-particle" state (it could hardly be called a vacuum) and endeavoured to show that the rapidly varying metric would have created pairs of particles with an approximately thermal distribution. However it is not clear whether these "no-particle" initial conditions have any physical significance or, indeed, whether one should try to start off the universe with only gravity and nothing else. This would seem unfair discrimination in favour of gravitons over other particles.

I would like to suggest a different approach to the problem of the boundary conditions of the universe which is based on the so-called Euclidean approach to quantum gravity. In this one evaluates the path integral over all positive definite (Euclidean) metrics with certain boundary conditions. This improves the convergence of the path integral and enables one to handle topologically non-trivial manifolds which cannot really be done if one restricts the path integral to metrics of Lorentzian signature. It also improves the situation with regard to singularities because these will still exist in the complexified spacetime manifold but they can be avoided by the section (Euclidean section) of the manifold on which the metric is real and positive definite. An example of this is again provided by black holes. The partition function Z_I for the canonical ensemble at a temperature T is defined by a path integral over all metrics which are asymptotically flat in the spatial dimension and which are periodic with a period $\beta = T^{-1}$ in the Euclidean time direction. The Schwarzschild solution is normally given in the form.

$$d s^2 = -\left(1 - \frac{2M}{r}\right) d t^2 + \left(1 - \frac{2M}{r}\right)^{-1} d r^2 + r^2 d \Omega^2$$

Putting $\tau = -i t$ one obtains a positive definite metric for $r > 2M$. The apparent singularity at $r = 2M$ can be removed by defining a new radial coordinate

$$x = 4M \left(1 - \frac{2M}{r}\right)^{1/2} .$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

The metric then becomes

$$d s^2 = x^2 (d \tau/4 M)^2 + (r/2 M)^4 d x^2 + r^2 d \Omega^2$$

The apparent singularity at $x = 0$, $r = 2 M$ is like the singularity at the origin of the polar coordinates provided that $\tau/4M$ is regarded as an angular variable and is identified with a period 2π, i.e. τ is identified with a period $8 \pi M$. The resultant manifold, the Euclidean section of the Schwarzschild solution, is asymptotically flat and non-singular because it does not contain any points with $r < 2M$. The curvature singularity at $r = 0$ does not lie on the Euclidean section. Thus one does not need to specify the boundary values of physical fields at the singularity.

One might ask whether one could find such a singularity free Euclidean section for the Friedmann-Robertson-Walker metrics. The answer is that one can, at least for the spatially closed solutions. For example, the Tolman solution for $p = 1/3 \mu$ is

$$d s^2 = -d t^2 + (a^2 - t^2) d \Omega^2_3 \ ,$$

where $d \Omega^2_3$ is the metric on a three sphere. The quantity 'a' is the radius of the universe at the time of maximum expansion, $t = 0$. The Big Bang is at $t = - a$ and the Big Crunch is at $t = + a$. If one puts $\tau = i t$, one obtains a complete singularity free positive definite metric which, as shown in Figure 4, has two asymptotically Euclidean regions connected by a throat of radius 'a'.

This metric avoids the problem of having to specify boundary conditions at a singularity but one would have to explain why the positive definite metric should be asymptotically Euclidean in two regions and why the throat between them should be so large: the radius of 'a' of the throat is at least 10^{60} Planck lengths. One possible reason might be that this metric is in fact a constant time section of a five dimensional black hole. The entropy of this five dimensional black hole would be proportional to a^3. One might be able to interpret this entropy as the logarithm of the volume in space of all metrics occupied by metrics near the given metric. In this case the most probable value for 'a' would be infinite, i.e., one would have a $k = 0$ parabolic universe. Within the limits of observational accuracy, this is indeed what we have.

This positive definite metric does not, however, explain why the universe should have started off in thermal equilibrium at a high temperature, if indeed it did. To understand this I think that it is essential to consider the large fluctuations of the metric and of the topology of

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

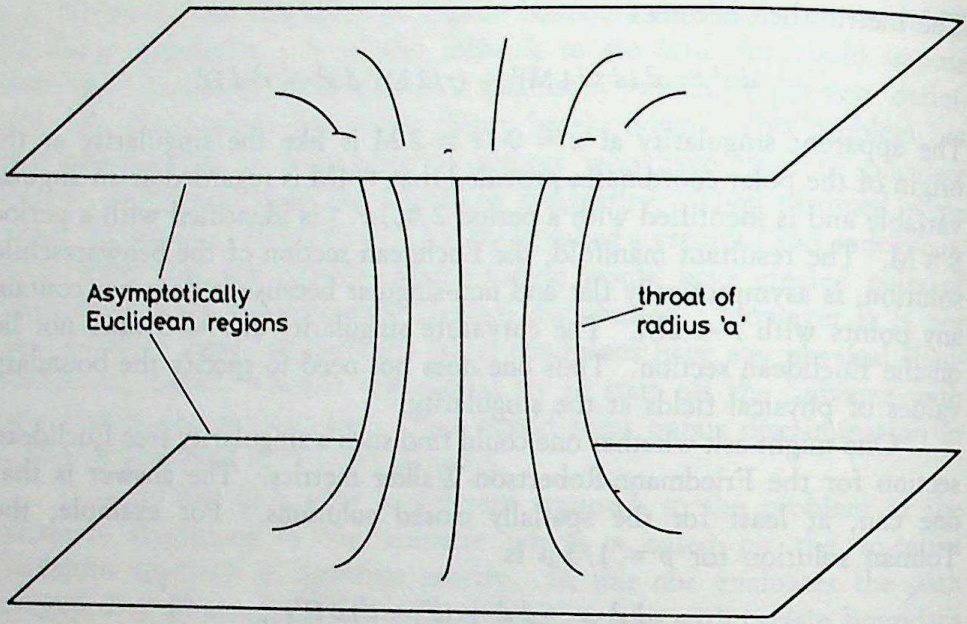


FIG. 4. A singularity-free positive definite metric which has two asymptotically Euclidean regions connected by a throat of radius, a.

spacetime that one would expect to occur on the order of the Planck length or less. One might imagine that spacetime would display a "foam-like" or even a fractal structure on these length scales but that it would appear smooth and nearly flat on larger length scales. In order to study this it is helpful to consider the volume canonical ensemble which is defined by a path integral over all compact positive definite metrics with an action that includes a cosmological term which is interpreted as a Lagrange multiplier for the four-volume, just as the inverse temperature is a Lagrange multiplier for the energy in the ordinary canonical ensemble. By evaluating the path integral over compact metrics, one eliminates one of the two parts of physics, the boundary conditions. There ought to be something very special about the boundary conditions of the universe and what can be more special than the condition that there is no boundary.

It seems reasonable to suppose that the path integral is dominated by metrics which are very complicated on Planck length scales but which appear smooth on a larger length scale and that the number of gravitational states with a four-volume V goes up steeply with increasing V. Any non-zero cosmological constant which arose from the vacuum fluctuations of the various fields or from non-zero Higgs potentials could be absorbed

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

into a shift in the Lagrange multiplier and a slight change in the small scale structure. Thus one might be able to explain both why the apparent cosmological constant is zero or nearly zero and why the universe is so near to a parabolic $k = 0$ model. The argument in the latter case would be that there were indeed states of the gravitational field which corresponded to universes which were only a few Planck lengths across, but that the number of gravitational states which they represented would be much smaller than for large universes. Normally, the value of the cosmological constant is regarded as belonging to the equation-of-motion part of the physics whereas the spatial flatness or otherwise of the universe is supposed to be a boundary condition. The volume canonical ensemble offers a possible way in which they might be related.

The complicated topology of the spacetime foam will mean that spacetime will not have the normal causal properties such as propagation only within the light cone and global hyperbolicity. These acausal effects will not be noticeable at low energies but they will be important near the Big Bang. One can think of the foam as being made up of virtual black holes which appear and disappear in about a Planck time, 10^{-43} seconds. Particles with normal energies will have wavelengths much longer than the Planck length and will have very little chance of falling into these virtual black holes. However, very high energy particles will be able to fall into the virtual black holes and will be re-emitted as particles of other species. This will mean that the universe will necessarily be in thermal equilibrium at the Planck time. The acausal properties of the spacetime foam might explain why the universe started off spatially homogeneous and isotropic even though different regions were not in causal contact according to the classical metric. This might also be a way of avoiding the production of too many monopoles, a problem that does not seem to have any satisfactory answer in conventional theories.

To summarise, because of quantum fluctuations of the causal structure of spacetime, it is not sufficient to impose boundary conditions at the Big Bang singularity, even if they could be well-defined there. To overcome this difficulty, I would advocate the Euclidean approach to quantum gravity in which the path integral is evaluated over positive definite metrics. If one uses compact metrics, one gets rid of the need to specify boundary conditions for the universe. Hopefully this approach would explain:

1. Why the apparent cosmological constant is zero;
2. Why the universe is spatially flat;
3. Why the universe was in thermal equilibrium.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PEEBLES

If the universe were in thermal equilibrium at the Planck time (strict Gibbs distribution ignoring gravitational energy) the thermal fluctuations could be disastrously large.

HAWKING

It seems that the universe was in *local* thermal equilibrium near the Big Bang, but I agree that if you assume that it was in thermal equilibrium over larger scales, you get too much large-scale inhomogeneity now. This is another of the problems that one gets into if one tries to impose a boundary condition at the initial singularity. On the other hand, if one imposes boundary conditions on the Euclidean section, one might expect the universe to be smooth on the Euclidean section. In the case of the $K = +1$ Tolman model, this would say that the universe would be smooth at the time of maximum expansion. The problem then would be to explain why it is slightly inhomogeneous now.

SILK

The most likely form of fluctuations at the quantum gravity epoch may be constant curvature fluctuations. These cannot have a large amplitude because of large-scale anisotropy constraints. Is it possible that, even on sub-horizon scales, black hole formation could be suppressed as a consequence of this?

HAWKING

In order to show that one would expect to form a large number of black holes of a few Planck masses, I assume only that one has thermal fluctuations within the particle horizon. As you and Peebles point out, the fluctuations on scales much larger than the horizon must be less than would be expected for thermal equilibrium.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

SCIAMA

I understand that not all curved spaces of indefinite metric can be analytically continued into space with positive definite metric. Are you depending essentially on the fact that it can be done for a Robertson-Walker metric?

HAWKING

It is true that only rather special metrics, like the Schwarzschild and Robertson-Walker metrics, admit a Euclidean section on which the metric is real and positive definite. However, the idea is that a path integral over *all* positive definite metrics is equivalent to a path integral over all Lorentzian metrics. It is similar to what is done in ordinary field theory: one integrates over all fields that are real in Euclidean space and says that it is equivalent to integrating over fields that are real in Minkowski space.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PHYSICAL REVIEW D VOLUME 28, NUMBER 12 15 DECEMBER 1983

Wave function of the Universe

J. B. Hartle
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637
and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

S. W. Hawking
Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge, England
and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106
(Received 29 July 1983)

The quantum state of a spatially closed universe can be described by a wave function which is a functional on the geometries of compact three-manifolds and on the values of the matter fields on these manifolds. The wave function obeys the Wheeler-DeWitt second-order functional differential equation. We put forward a proposal for the wave function of the "ground state" or state of minimum excitation: the ground-state amplitude for a three-geometry is given by a path integral over all compact positive-definite four-geometries which have the three-geometry as a boundary. The requirement that the Hamiltonian be Hermitian then defines the boundary conditions for the Wheeler-DeWitt equation and the spectrum of possible excited states. To illustrate the above, we calculate the ground and excited states in a simple minisuperspace model in which the scale factor is the only gravitational degree of freedom, a conformally invariant scalar field is the only matter degree of freedom and $\Lambda > 0$. The ground state corresponds to de Sitter space in the classical limit. There are excited states which represent universes which expand from zero volume, reach a maximum size, and then recollapse but which have a finite (though very small) probability of tunneling through a potential barrier to a de Sitter-type state of continual expansion. The path-integral approach allows us to handle situations in which the topology of the three-manifold changes. We estimate the probability that the ground state in our minisuperspace model contains more than one connected component of the spacelike surface.

I. INTRODUCTION

In any attempt to apply quantum mechanics to the Universe as a whole the specification of the possible quantum-mechanical states which the Universe can occupy is of central importance. This specification determines the possible dynamical behavior of the Universe. Moreover, if the uniqueness of the present Universe is to find any explanation in quantum gravity it can only come from a restriction on the possible states available.

In quantum mechanics the state of a system is specified by giving its wave function on an appropriate configuration space. The possible wave functions can be constructed from the fundamental quantum-mechanical amplitude for a complete history of the system which may be regarded as the starting point for quantum theory.[1] For example, in the case of a single particle a history is a path $x(t)$ and the amplitude for a particular path is proportional to

$$\exp(iS[x(t)]) , \tag{1.1}$$

where $S[x(t)]$ is the classical action. From this basic amplitude, the amplitude for more restricted observations can be constructed by superposition. In particular, the amplitude that the particle, having been prepared in a certain way, is located at position x and nowhere else at time t is

$$\psi(x,t) = N \int_C \delta x(t) \exp(iS[x(t)]) . \tag{1.2}$$

Here, N is a normalizing factor and the sum is over a class of paths which intersect x at time t and which are weighted in a way that reflects the preparation of the system. $\psi(x,t)$ is the wave function for the state determined by this preparation. As an example, if the particle were previously localized at x' at time t' one would sum over all paths which start at x' at t' and end at x at t thereby obtaining the propagator $\langle x,t \,|\, x',t' \rangle$. The oscillatory integral in Eq. (1.2) is not well defined but can be made so by rotating the time to imaginary values.

An alternative way of calculating quantum dynamics is to use the Schrödinger equation,

$$i\partial\psi/\partial t = H\psi . \tag{1.3}$$

This follows from Eq. (1.2) by varying the end conditions on the path integral. For a particular state specified by a weighting of paths C, the path integral (1.2) may be looked upon as providing the boundary conditions for the solution of Eq. (1.3).

A state of particular interest in any quantum-mechanical theory is the ground state, or state of minimum excitation. This is naturally defined by the path integral, made definite by a rotation to Euclidean time, over the class of paths which have vanishing action in the far past. Thus, for the ground state at $t=0$ one would write

$$\psi_0(x,0) = N \int \delta x(\tau) \exp(-I[x(\tau)]) , \tag{1.4}$$

where $I[x(\tau)]$ is the Euclidean action obtained from S by

 ©1983 The American Physical Society

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

sending $t \to -i\tau$ and adjusting the sign so that it is positive.

In cases where there is a well-defined time and a corresponding time-independent Hamiltonian, this definition of ground state coincides with the lowest eigenfunction of the Hamiltonian. To see this specialize the path-integral expression for the propagator $\langle x,t \mid x',t' \rangle$ to $t=0$ and $x'=0$ and insert a complete set of energy eigenstates between the initial and final state. One has

$$\langle x,0 \mid 0,t' \rangle = \sum_n \psi_n(x)\bar{\psi}_n(0)\exp(iE_n t') = \int \delta x(t)\exp(iS[x(t)]) , \tag{1.5}$$

where $\psi_n(x)$ are the time-independent energy eigenfunctions. Rotate $t' \to -i\tau'$ in (1.5) and take the limit as $\tau' \to -\infty$. In the sum only the lowest eigenfunction (normalized to zero energy) survives. The path integral becomes the path integral on the right of (1.4) so that the equality is demonstrated.

The case of quantum fields is a straightforward generalization of quantum particle mechanics. The wave function is a functional of the field configuration on a spacelike surface of constant time, $\Psi=\Psi[\phi(\vec{x}),t]$. The functional Ψ gives the amplitude that a particular field distribution $\phi(\vec{x})$ occurs on this spacelike surface. The rest of the formalism is similarly generalized. For example, for the ground-state wave functional one has

$$\Psi_0[\phi(\vec{x}),0]=N \int \delta\phi(x)\exp(-I[\phi(x)]) , \tag{1.6}$$

where the integral is over all Euclidean field configurations for $\tau<0$ which match $\phi(\vec{x})$ on the surface $\tau=0$ and leave the action finite at Euclidean infinity.

In the case of quantum gravity new features enter. For definiteness and simplicity we shall restrict our attention throughout this paper to spatially closed universes. For these there is no well-defined intrinsic measure of the location of a spacelike surface in the spacetime beyond that contained in the intrinsic or extrinsic geometry of the surface itself. One therefore labels the wave function by the three-metric h_{ij} writing $\Psi=\Psi[h_{ij}]$. Quantum dyanmics is supplied by the functional integral

$$\Psi[h_{ij}]=N \int_C \delta g(x)\exp(iS_E[g]) . \tag{1.7}$$

S_E is the classical action for gravity including a cosmological constant Λ and the functional integral is over all four-geometries with a spacelike boundary on which the induced metric is h_{ij} and which to the past of that surface satisfy some appropriate condition to define the state. In particular for the amplitude to go from a three-geometry h'_{ij} on an initial spacelike surface to a three-geometry h''_{ij} on a final spacelike surface is

$$\langle h''_{ij} \mid h'_{ij} \rangle = \int \delta g \exp(iS_E[g]) , \tag{1.8}$$

where the sum is over all four-geometries which match h'_{ij} on the initial surface and h''_{ij} on the final surface. Here one clearly sees that one cannot specify time in these states. The proper time between the surfaces depends on the four-geometries in the sum.

As in the mechanics of a particle the functional integral (1.7) implies a differential equation on the wave function. This is the Wheeler-DeWitt equation[2] which we shall derive from this point of view in Sec. II. With a simple choice of factor ordering it is

$$\left[-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}} - {}^3R(h)h^{1/2} + 2\Lambda h^{1/2} \right] \Psi[h_{ij}]=0 , \tag{1.9}$$

where G_{ijkl} is the metric on superspace,

$$G_{ijkl}=\tfrac{1}{2}h^{-1/2}(h_{ik}h_{jl}+h_{il}h_{jk}-h_{ij}h_{kl}) \tag{1.10}$$

and ${}^3R$ is the scalar curvature of the intrinsic geometry of the three-surface. The problem of specifying cosmological states is the same as specifying boundary conditions for the solution of the Wheeler-DeWitt equation. A natural first question to ask is what boundary conditions specify the ground state?

In the quantum mechanics of closed universes we do not expect to find a notion of ground state as a state of lowest energy. There is no natural definition of energy for a closed universe just as there is no independent standard of time. Indeed in a certain sense the total energy for a closed universe is always zero—the gravitational energy canceling the matter energy. It is still reasonable, however, to expect to be able to define a state of minimum excitation corresponding to the classical notion of a geometry of high symmetry. This paper contains a proposal for the definition of such a ground-state wave function for closed universes. The proposal is to extend to gravity the Euclidean-functional-integral construction of nonrelativistic quantum mechanics and field theory [Eqs. (1.4) and (1.6)]. Thus, we write for the ground-state wave function

$$\Psi_0[h_{ij}]=N \int \delta g \exp(-I_E[g]) , \tag{1.11}$$

where I_E is the Euclidean action for gravity including a cosmological constant Λ. The Euclidean four-geometries summed over must have a boundary on which the induced metric is h_{ij}. The remaining specification of the class of geometries which are summed over determines the ground state. Our proposal is that the sum should be over compact geometries. This means that the Universe does not have any boundaries in space or time (at least in the Euclidean regime) (cf. Ref. 3). There is thus no problem of boundary conditions. One can interpret the functional integral over all compact four-geometries bounded by a given three-geometry as giving the amplitude for that three-geometry to arise from a zero three-geometry, i.e., a single point. In other words, the ground state is the amplitude for the Universe to appear from nothing.[4] In the following we shall elaborate on this construction and show in simple models that it indeed supplies reasonable wave functions for a state of minimum excitation.

The specification of the ground-state wave function is a constraint on the other states allowed in the theory. They must be such, for example, as to make the Wheeler-DeWitt equation Hermitian in an appropriate norm. In analogy with ordinary quantum mechanics one would expect to be able to use these constraints to extrapolate the boundary conditions which determine the excited states of

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the theory from those fixed for the ground state by Eq. (1.7). Thus, one can in principle determine all the allowed cosmological states.

The wave functions which result from this specification will not vanish on the singular, zero-volume three-geometries which correspond to the big-bang singularity. This is analogous to the behavior of the wave function of the electron in the hydrogen atom. In a classical treatment, the situation in which the electron is at the proton is singular. However, in a quantum-mechanical treatment the wave function in a state of zero angular momentum is finite and nonzero at the proton. This does not cause any problems in the case of the hydrogen atom. In the case of the Universe we would interpret the fact that the wave function can be finite and nonzero at the zero three-geometry as allowing the possibility of topological fluctuations of the three-geometry. This will be discussed further in Sec. VIII.

After a general discussion of this proposal for the ground-state wave function we shall implement it in a minisuperspace model. The geometrical degrees of freedom in the model are restricted to spatially homogeneous, isotropic, closed universes with S^3 topology, the matter degrees of freedom to a single, homogeneous, conformally invariant scalar field and the cosmological constant is assumed to be positive. A semiclassical evaluation of the functional integral for the ground-state wave function shows that it indeed does possess characteristics appropriate to a "state of minimum excitation."

Extrapolating the boundary conditions which allow the ground state to be extracted from the Wheeler-DeWitt equation, we are able to go further and identify the wave functions in the minisuperspace models corresponding to excited states of the matter field. These wave functions display some interesting features. One has a complete spectrum of excited states which show that a closed universe similar to our own and possessed of a cosmological constant can escape the big crunch and tunnel through to an eternal de Sitter expansion. We are able to calculate the probability for this transition.

In addition to the excited states we make a proposal for the amplitudes that the ground-state three-geometry consists of disconnected three-spheres thus giving a meaning to a gravitational state possessing different topologies.

Our conclusion will be that the Euclidean-functional-integral prescription (1.7) does single out a reasonable candidate for the ground-state wave function for cosmology which when coupled with the Wheeler-DeWitt equation yields a basis for constructing quantum cosmologies.

II. QUANTUM GRAVITY

In this section we shall review the basic principles and machinery of quantum gravity with which we shall explore the wave functions for closed universes. For simplicity we shall represent the matter degrees of freedom by a single scalar field ϕ, more realistic cases being straightforward generalizations. We shall approach this review from the functional-integral point of view although we shall arrive at many canonical results.[5] None of these are new and for different approaches to the same ends the reader is referred to the standard literature.[6]

A. Wave functions

Our starting point is the quantum-mechanical amplitude for the occurrence of a given spacetime and a given field history. This is

$$\exp(iS[g,\phi]) , \tag{2.1}$$

where $S[g,\phi]$ is the total classical action for gravity coupled to a scalar field. We are envisaging here a fixed manifold although there is no real reason that amplitudes for different manifolds may not be considered provided a rule is given for their relative phases. Just as the interesting observations of a particle are not typically its entire history but rather observations of position at different times, so also the interesting quantum-mechanical questions for gravity correspond to observations of spacetime and field on different spacelike surfaces. Following the general rules of quantum mechanics the amplitudes for these more restricted sets of observations are obtained from (2.1) by summing over the unobserved quantities.

It is easy to understand what is meant by fixing the field on a given spacelike surface. What is meant by fixing the four-geometry is less obvious. Consider all four-geometries in which a given spacelike surface occurs but whose form is free to vary off the surface. By an appropriate choice of gauge near the surface (e.g., Gaussian normal coordinates) all these four-geometries can be expressed so that the only freedom in the four-metric is the specification of the three-metric h_{ij} in the surface. Specifying the three-metric is therefore what we mean by fixing the four-geometry on a spacelike surface. The situation is not unlike gauge theories. There a history is specified by a vector potential $A_\mu(x)$ but by an appropriate gauge transformation $A_0(x)$ can be made to vanish so that the field on a surface can be completely specified by the $A_i(x)$.

As an example of the quantum-mechanical superposition principle the amplitude for the three-geometry and field to be fixed on two spacelike surfaces is

$$\langle h_{ij}'',\phi'' \mid h_{ij}',\phi' \rangle = \int \delta g\, \delta\phi \exp(iS[g,\phi]) , \tag{2.2}$$

where the integral is over all four-geometries and field configurations which match the given values on the two spacelike surfaces. This is the natural analog of the propagator $\langle x'',t' \mid x',t' \rangle$ in the quantum mechanics of a single particle. We note again that the proper time between the two surfaces is not specified. Rather it is summed over in the sense that the separation between the surfaces depends on the four-geometry being summed over. It is not that one could not ask for the amplitude to have the three-geometry and field fixed on two surfaces *and* the proper time between them. One could. Such an amplitude, however, would not correspond to fixing observations on just two surfaces but rather would involve a set of intermediate observations to determine the time. It would therefore not be the natural analog of the propagator.

Wave functions Ψ are defined by

$$\Psi[h_{ij},\phi] = \int_C \delta g\, \delta\phi \exp(iS[g,\phi]) . \tag{2.3}$$

The sum is over a class C of spacetimes with a compact boundary on which the induced metric is h_{ij} and field configurations which match ϕ on the boundary. The

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

remaining specification of the class C is the specification of the state.

If the Universe is in a quantum state specified by a wave function Ψ then that wave function describes the correlations between observables to be expected in that state. For example, in the semiclassical wave function describing a universe like our own, one would expect Ψ to be large when ϕ is big and the spatial volume is small, large when ϕ is small and the spatial volume is big, and small when these quantities are oppositely correlated. This is the only interpretative structure we shall propose or need.

B. Wheeler-DeWitt equation

A differential equation for Ψ can be derived by varying the end conditions on the path integral (2.3) which defines it. To carry out this derivation first recall that the gravitational action appropriate to keeping the three-geometry fixed on a boundary is

$$l^2 S_E = 2\int_{\partial M} d^3x\, h^{1/2}K + \int_M d^4x(-g)^{1/2}(R-2\Lambda)\,. \tag{2.4}$$

The second term is integrated over spacetime and the first over its boundary. K is the trace of the extrinsic curvature K_{ij} of the boundary three-surface. If its unit normal is n^i, $K_{ij} = -\nabla_i n_j$ in the usual Lorentzian convention. l is the Planck length $(16\pi G)^{1/2}$ in the units with $\hbar = c = 1$ we use throughout. Introduce coordinates so that the boundary is a constant t surface and write the metric in the standard $3+1$ decomposition:

$$ds^2 = -(N^2 - N_i N^i)dt^2 + 2N_i dx^i dt + h_{ij}dx^i dx^j\,. \tag{2.5}$$

The action (2.4) becomes

$$l^2 S_E = \int d^4x\, h^{1/2} N[K_{ij}K^{ij} - K^2 + {}^3R(h) - 2\Lambda]\,, \tag{2.6}$$

where explicitly

$$K_{ij} = \frac{1}{N}\left[-\frac{1}{2}\frac{\partial h_{ij}}{\partial t} + N_{(i\,|\,j)}\right] \tag{2.7}$$

and a stroke and ${}^3R$ denote the covariant derivative and scalar curvature constructed from the three-metric h_{ij}. The matter action S_M can similarly be expressed as a function of N, N_i, h_{ij}, and the matter field.

The functional integral defining the wave function contains an integral over N. By varying N at the surface we push it forward or backward in time. Since the wave function does not depend on time we must have

$$0 = \int \delta g\, \delta\phi \left[\frac{\delta S}{\delta N}\right] \exp(iS[g,\phi])\,. \tag{2.8}$$

More precisely, the value of the integral (2.3) should be left unchanged by an infinitesimal translation of the integration variable N. If the measure is invariant under translation this leads to (2.8). If it is not, there will be in addition a divergent contribution to the relation which must be suitably regulated to zero or cancel divergences arising from the calculation of the right-hand side of (2.8).

Classically the field equation $H \equiv \delta S/\delta N = 0$ is the Hamiltonian constraint for general relativity. It is

$$H = h^{1/2}(K^2 - K_{ij}K^{ij} + {}^3R - 2\Lambda - l^2 T_{nn}) = 0\,, \tag{2.9}$$

where T_{nn} is the stress-energy tensor of the matter field projected in the direction normal to the surface. Equation (2.8) shows how $H=0$ is enforced as an operator identity for the wave function. More explicitly one can note that the K_{ij} involve only first-time derivatives of the h_{ij} and therefore may be completely expressed in terms of the momenta π_{ij} conjugate to the h_{ij} which follow from the Lagrangian in (2.6):

$$\pi_{ij} = -h^{1/2}(K_{ij} - h_{ij}K)\,. \tag{2.10}$$

In a similar manner the energy of the matter field can be expressed in terms of the momentum conjugate to the field π_ϕ and the field itself. Equation (2.8) thus implies the operator identity $H(\pi_{ij}, h_{ij}, \pi_\phi, \phi)\Psi = 0$ with the replacements

$$\pi^{ij} = -i\frac{\delta}{\delta h_{ij}}\,,\quad \pi_\phi = -i\frac{\delta}{\delta\phi}\,. \tag{2.11}$$

These replacements may be viewed as arising directly from the functional integral, e.g., from the observation that when the time derivatives in the exponent are written in differenced form

$$-i\frac{\delta}{\delta h_{ij}}\int \delta g\, \delta\phi\, e^{iS} = \int \delta g\, \delta\phi\, \pi^{ij} e^{iS}\,. \tag{2.12}$$

Alternatively, they are the standard representation of the canonical commuation relations of h_{ij} and π^{ij}.

In translating a classical equation like $\delta S/\delta N = 0$ into an operator identity there is always the question of factor ordering. This will not be important for us so making a convenient choice we obtain

$$\left\{-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}} + h^{1/2}\left[-{}^3R(h) + 2\Lambda + l^2 T_{nn}\left(-i\frac{\delta}{\delta\phi}, \phi\right)\right]\right\} \times \Psi[h_{ij},\phi] = 0\,. \tag{2.13}$$

This is the Wheeler-DeWitt equation which wave functions for closed universes must satisfy. There are also the other constraints of the classical theory, but the operator versions of these express the gauge invariance of the wave function rather than any dynamical information.[6]

We should emphasize that the ground-state wave function constructed by a Euclidean functional-integral prescription [(Eq. (1.11)] will satisfy the Wheeler-DeWitt equation in the form (2.13). Indeed, this can be demonstrated explicitly by repeating the steps in the above demonstration starting with the Euclidean functional integral.

C. Boundary conditions

The quantity G_{ijkl} can be viewed as a metric on superspace—the space of all three-geometries (no connection with supersymmetry). It has signature

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$(-,+,+,+,+,+)$ and the Wheeler-DeWitt equation is therefore a "hyperbolic" equation on superspace. It would be natural, therefore, to expect to impose boundary conditions on two "spacelike surfaces" in superspace. A convenient choice for the timelike direction is $h^{1/2}$ and we therefore expect to impose boundary conditions at the upper and lower limits of the range of $h^{1/2}$. The upper limit is infinity. The lower limit is zero because if h_{ij} is positive definite or degenerate, $h^{1/2} \geq 0$. Positive-definite metrics are everywhere spacelike surfaces; degenerate metrics may signal topology change. Summarizing the remaining functions of h_{ij} by the conformal metric $\tilde{h}_{ij} = h_{ij}/h^{1/3}$ we may write an important boundary condition on Ψ as

$$\Psi[\tilde{h}_{ij}, h^{1/2}, \phi] = 0, \quad h^{1/2} < 0 . \tag{2.14}$$

Because $h^{1/2}$ has a semidefinite range it is for many purposes convenient to introduce a representation in which $h^{1/2}$ is replaced by its canonically conjugate variable $-\frac{4}{3}Kl^{-2}$ which has an infinite range. The advantages of this representation have been extensively discussed.[7] In the case of pure gravity since $-\frac{4}{3}Kl^{-2}$ and $h^{1/2}$ are conjugate, we can write for the transformation to the representation where $\tilde{h}_{ij}$ and K are definite

$$\Phi[\tilde{h}_{ij}, K] = \int_0^\infty \delta h^{1/2} \exp\left[-i\tfrac{4}{3}l^{-2}\int d^3x\, h^{1/2}K\right]\Psi[h_{ij}] \tag{2.15}$$

and inversely,

$$\Psi[h_{ij}] = \int_{-\infty}^{+\infty} \delta K \exp\left[+i\tfrac{4}{3}l^{-2}\int d^3x\, h^{1/2}K\right]\Phi[\tilde{h}_{ij}, K] . \tag{2.16}$$

In each case the functional integrals are over the values of $h^{1/2}$ or K at each point of the spacelike hypersurface and we have indicated limits of integration.

The condition (2.14) implies through (2.15) that $\Phi[\tilde{h}_{ij}, K]$ is analytic in the lower-half K plane. The contour in (2.16) can thus be distorted into the lower-half K plane. Conversely, if we are given $\Phi[\tilde{h}_{ij}, K]$ we can reconstruct the wave function Ψ which satisfies the boundary condition (2.14) by carrying out the integration in (2.16) over a contour which lies below any singularities of $\Phi[\tilde{h}_{ij}, K]$ in K.

In the presence of matter K and $\tilde{h}_{ij}$ remain convenient labels for the wave functional provided the labels for the matter-field amplitudes $\tilde{\phi}$ are chosen so that a multiple of K is canonically conjugate to $h^{1/2}$. In cases where the matter-field action itself involves the scalar curvature this means that the label $\tilde{\phi}$ will be the field amplitude rescaled by some power of $h^{1/2}$. For example, in the case of a conformally invariant scalar field the appropriate label is $\tilde{\phi} = \phi h^{1/6}$. With this understanding we can write for the functionals

$$\Psi = \Psi[h_{ij}, \tilde{\phi}], \quad \Phi = \Phi[\tilde{h}_{ij}, K, \tilde{\phi}] \tag{2.17}$$

and the transformation formulas (2.15) and (2.16) remain unchanged.

D. Hermiticity

The introduction of wave functions as functional integrals [Eq. (2.3)] allows the definition of a scalar product with a simple geometric interpretation in terms of sums over spacetime histories. Consider a wave function Ψ defined by the integral

$$\Psi[h_{ij}, \phi] = N \int_C \delta g\, \delta\phi \exp(iS[g, \phi]) , \tag{2.18}$$

over a class of four-geometries and fields C, and a second wave function Ψ' defined by a similar sum over a class C'. The scalar product

$$(\Psi', \Psi) = \int \delta h\, \delta\phi\, \bar{\Psi}'[h_{ij}, \phi]\Psi[h_{ij}, \phi] \tag{2.19}$$

has the geometric interpretation of a sum over all histories

$$(\Psi', \Psi) = \bar{N}'N \int \delta g\, \delta\phi \exp(iS[g, \phi]) , \tag{2.20}$$

where the sum is over histories which lie in class C to the past of the surface and in the time reversed of class C' to its future.

The scalar product (2.19) is not the product that would be required by canonical theory to define the Hilbert space of physical states. That would presumably involve integration over a hypersurface in the space of all three-geometries rather than over the whole space as in (2.19). Rather, Eq. (2.19) is a mathematical construction made natural by the functional-integral formulation of quantum gravity.

In gravity we expect the field equations to be satisfied as identities. An extension of the argument leading to Eq. (2.8) will give

$$\int \delta g\, \delta\phi\, H(x) \exp(iS[g, \phi]) = 0 \tag{2.21}$$

for any class of geometries summed over and for any intermediate spacelike surface on which $H(x)$ is evaluated. Equation (2.21) can be evaluated for the particular sum which enters Eq. (2.20). $H(x)$ can be interpreted in the scalar product as an operator acting on either Ψ' or Ψ. Thus,

$$(H\Psi', \Psi) = (\Psi', H\Psi) = 0 . \tag{2.22}$$

The Wheeler-DeWitt operator must therefore be Hermitian in the scalar product (2.19).

Since the Wheeler-DeWitt operator is a second-order functional-differential operator, the requirement of Hermiticity will essentially be a requirement that certain surface terms on the boundary of the space of three-metrics vanish and, in particular, at $h^{1/2} = 0$ and $h^{1/2} = \infty$. As in ordinary quantum mechanics these conditions will prove useful in providing boundary conditions for the solution of the equation.

III. GROUND-STATE WAVE FUNCTION

In this section, we shall put forward in detail our proposal for the ground-state wave function for closed cosmologies. The wave function depends on the topology and the three-metric of the spacelike surface and on the values of the matter field on the surface. For simplicity we shall begin by considering only S^3 topology. Other

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

possibilities will be considered in Sec. VIII.

As discussed in the Introduction, the ground-state wave function is to be constructed as a functional integral of the form

$$\Psi_0[h_{ij},\phi]=N\int \delta g\,\delta\phi\exp(-I[g,\phi])\,, \tag{3.1}$$

where I is the total Euclidean action and the integral is over an appropriate class of Euclidean four-geometries with compact boundary on which the induced metric is h_{ij} and an appropriate class of Euclidean field configurations which match the value given on the boundary. To complete the definition of the ground-state wave function we need to give the class of geometries and fields to be summed over. Our proposal is that the geometries should be compact and that the fields should be regular on these geometries. In the case of a positive cosmological constant Λ any regular Euclidean solution of the field equations is necessarily compact.[8] In particular, the solution of greatest symmetry is the four-sphere of radius $3/\Lambda$, whose metric we write as

$$ds^2=(\sigma/H)^2(d\theta^2+\sin^2\theta\,d\Omega_3{}^2)\,, \tag{3.2}$$

where $d\Omega_3{}^2$ is the metric on the three-sphere. $H^2=\sigma^2\Lambda/3$ and we have introduced the normalization factor $\sigma^2=l^2/24\pi^2$ for later convenience. Thus, it is clear that compact four-geometries are the only reasonable candidates for the class to be summed over when $\Lambda>0$.

If Λ is zero or negative there are noncompact solutions of the field equations. The solutions of greatest symmetry are Euclidean space ($\Lambda=0$) with

$$ds^2=\sigma^2(d\theta^2+\theta^2\,d\Omega_3{}^2) \tag{3.3}$$

and Euclidean anti–de Sitter space ($\Lambda<0$) with

$$ds^2=(\sigma/H)^2(d\theta^2+\sinh^2\theta\,d\Omega_3{}^2)\,. \tag{3.4}$$

One might therefore feel that the ground state for $\Lambda\le 0$ should be defined by a functional integral over geometries which are asymptotically Euclidean or asymptotically anti–de Sitter. This is indeed appropriate to defining the ground state for scattering problems where one is interested in particles which propagate in from infinity and then out to infinity again.[9] However, in the case of cosmology, one is interested in measurements that are carried out in the interior of the spacetime, whether or not the interior points are connected to some infinite regions does not matter. If one were to use asymptotically Euclidean or anti–de Sitter four-geometries in the functional integral that defines the ground state one could not exclude a contribution from four-geometries that consisted of two disconnected pieces, one of which was compact with the three-geometry as boundary and the other of which was asymptotically Euclidean or anti–de Sitter with no interior boundary. Such disconnected geometries would in fact give the dominant contribution to the ground-state wave function. Thus, one would effectively be back with the prescription given above.

The ground-state wave function obtained by summing over compact four-geometries diverges for large three-geometries in the cases $\Lambda\le 0$ and the wave function cannot be normalized. This is because the Λ in the action damps large four-geometries when $\Lambda>0$, but it enhances them when $\Lambda<0$. We shall therefore consider only the case $\Lambda>0$ in this paper and shall regard $\Lambda=0$ as a limiting case of $\Lambda>0$.

An equivalent way of describing the ground state is to specify its wave function in the $\tilde\phi,\tilde h_{ij},K$ representation. Here too it can be constructed as a functional integral:

$$\Phi_0[\tilde h_{ij},K,\tilde\phi]=N\int \delta g\,\delta\phi\exp(-I^K[g,\phi])\,. \tag{3.5}$$

The sum is over the same class of fields and geometries as before except that now $\tilde\phi$, $\tilde h_{ij}$, and K are fixed on the boundary rather than ϕ and h_{ij}. The action I^K is therefore the Euclidean action appropriate to holding $\tilde\phi$, $\tilde h_{ij}$, and K fixed on a boundary. It is a sum of the appropriate pure gravitational action which up to an additive constant is

$$l^2I^K_E[g]=-\tfrac{2}{3}\int_{\partial M} d^3x\,h^{1/2}K-\int_M d^4x\,g^{1/2}(R-2\Lambda) \tag{3.6}$$

and a contribution from the matter. The latter is well illustrated by the action of a single conformally invariant scalar field, an example which we shall use exclusively in the rest of this paper. We have

$$I^K_M[g,\phi]=\tfrac{1}{2}\int_M d^4x\,g^{1/2}[(\nabla\phi)^2+\tfrac{1}{6}R\phi^2]\,. \tag{3.7}$$

These actions differ from the more familiar ones in which ϕ and h_{ij} are fixed only in having different surface terms. Indeed, these surface terms are just those required to ensure the equivalence of (3.1) and (3.5) as a consequence of the transformation formulas (2.15) and (2.16). In the case of the matter action of a conformally invariant scalar field with $\tilde\phi,\tilde h_{ij},K$ fixed the additional surface term conveniently cancels that required in the action when ϕ and h_{ij} are fixed.

It is important to recognize that the functional integral (3.5) does not yield the wave function at the Lorentzian value of K but rather at a Euclidean value of K. For the moment denote the Lorentzian value by K_L. If the hypersurfaces of interest were labeled by a time coordinate t in a coordinate system with zero shift [$N_i=0$ in Eq. (2.5)] then the rotation $t\to i\tau$ and the use of the traditional conventions $K_L=-\nabla\cdot n$ and $K=\nabla\cdot n$ will send $K_L\to -iK$. In terms of the Euclidean K the transformation formulas (2.15) and (2.16) can be rewritten to read

$$\Phi[\tilde h_{ij},K,\tilde\phi]=\int_0^\infty \delta h^{1/2}\exp\left[-\tfrac{4}{3}l^{-2}\int d^3x\,h^{1/2}K\right]\times\Psi[h_{ij},\tilde\phi]\,, \tag{3.8}$$

$$\Psi[h_{ij},\phi]=-\frac{1}{2\pi i}\int_C \delta K\exp\left[\tfrac{4}{3}l^{-2}\int d^3x\,h^{1/2}K\right]\times\Phi[\tilde h_{ij},K,\tilde\phi]\,, \tag{3.9}$$

where the contour C runs from $-i\infty$ to $+i\infty$. At the risk of some confusion we shall continue to use K in the remainder of this paper to denote the Euclidean K despite having used the same symbol in Secs. I and II for the Lorentzian quantity.

There is one advantage to constructing the ground-state wave function from the functional integral (3.5) rather than (3.1) and it is the following: the integral in Eq. (3.9)

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

will always yield a wave function $\Psi_0[h_{ij},\phi]$ which vanishes for $h^{1/2}<0$ if the contour C is chosen to the right of any singularities of $\Phi_0[\tilde{h}_{ij},K,\tilde{\phi}]$ in K provided Φ does not diverge too strongly in K. The boundary condition (2.14) is thus automatically enforced. This is a considerable advantage when the wave function is only evaluated approximately.

The Euclidean gravitational action [Eq. (3.6)] is not positive definite. The functional integrals in Eqs. (3.1) and (3.5) therefore require careful definition. One way of doing this is to break the integration up into an integral over conformal factors and over geometries in a given conformal equivalence class. By appropriate choice of the contour of integration of the conformal factor the integral can probably be made convergent. If this is the case a properly convergent functional integral can be constructed.

This then is our prescription for the ground state. In the following sections we shall derive some of its properties and demonstrate its reasonableness in a simple minisuperspace model.

IV. SEMICLASSICAL EXPECTATIONS

An important advantage of a functional-integral prescription for the ground-state wave function is that it yields the semiclassical approximation for that wave function directly. In this section, we shall examine the semiclassical approximation to the ground-state wave function defined in Sec. III. For simplicity we shall consider the case of pure gravity. The extension to include matter is straightforward.

The semiclassical approximation is obtained by evaluating the functional integral by the method of steepest descents. If there is only one stationary-phase point the semiclassical approximation is

$$\Psi_0[h_{ij}]=N\Delta^{-1/2}[h_{ij}]\exp(-I_{\rm cl}[h_{ij}])\ . \tag{4.1}$$

Here, $I_{\rm cl}$ is the Euclidean gravitational action evaluated at the stationary-phase point, that is, at that solution $g^{\rm cl}_{\mu\nu}$ of the Euclidean field equations

$$R_{\mu\nu}=\Lambda g_{\mu\nu}\ , \tag{4.2}$$

which induces the metric h_{ij} on the closed three-surface boundary and satisfies the asymptotic conditions discussed in Sec. III. $\Delta^{-1/2}$ is a combination of determinants of the wave operators defining the fluctuations about $g^{\rm cl}_{\mu\nu}$ including those contributed by the ghosts. We shall focus mainly on the exponent. For further information on Δ in the case without boundary see Ref. 10.

If there is more than one stationary-phase point, it is necessary to consider the contour of integration in the path integal more carefully in order to decide which gives the dominant contribution. In general this will be the stationary-phase point with the lowest value of ReI although it may not be if there are two stationary-phase points which correspond to four-metrics that are conformal to one another. We shall see an example of this in Sec. VI. The ground-state wave function is real. This means that if the stationary-phase points have complex values of the action, there will be equal contributions from stationary-phase points with complex-conjugate values of the action. If there is no four-geometry which is a stationary-phase point, the wave function will be zero in the semiclassical approximation.

The semiclassical approximation for Ψ_0 can also be obtained by first evaluating the semiclassical approximation to Φ_0 from the functional integral (3.5) and then evaluating the transformation integral (3.9) by steepest descents. This will be more convenient to do when the boundary conditions of fixing $\tilde{h}_{ij}$ and K yield a unique dominant stationary-phase solution to (4.2) but fixing h_{ij} does not.

One can fix the normalization constant N in (4.1) by the requirement

$$\int \delta h\, \bar{\Psi}_0[h_{ij}]\Psi_0[h_{ij}]=1\ . \tag{4.3}$$

As explained in Sec. II, one can interpret (4.3) geometrically as a path integral over all four-geometries which are compact on *both* sides of the three-surface with the metric h_{ij}. The semiclassical approximation to this path integral will thus be given by the action of the compact four-geometry without boundary which is the solution of the Einstein field equation. In the case of $\Lambda>0$ the solution with the most negative action is the four-sphere. Thus,

$$N^2=\exp\left[-\frac{2}{3H^2}\right]\ . \tag{4.4}$$

The semiclassical approximation for the wave function gives one considerable insight into the boundary conditions for the Wheeler-DeWitt equation, which are implied by the functional-integral prescription for the wave function. As discussed in Sec. II, these are naturally imposed on three-geometries of very large volumes and vanishing volumes.

Consider the limit of small three-volumes first. If the limiting three-geometry is such that it can be embedded in flat space then the classical solution to (4.2) when $\Lambda>0$ is the four-sphere and remains so as the three-geometry shrinks to zero. The action approaches zero. The value of the wave function is therefore controlled by the behavior of the determinants governing the fluctuations away from the classical solution. These fluctuations are to be computed about a vanishingly small region of a space of constant positive curvature. In this limit one can neglect the curvature and treat the fluctuations as about a region of flat space. The determinant can therefore be evaluated by considering its behavior under a constant conformal rescaling of the four-metric and the boundary three-metric. The change in the determinant under a change of scale is given by the value of the associated ζ function at zero argument.[11]

Regular four-geometries contain many hypersurfaces on which the three-volume vanishes. For example, consider the four-sphere of radius R embedded in a five-dimensional flat space. The three-surfaces which are the intersection of the four-sphere with surfaces of x^5 equals constant have a regular three-metric for $|x^5|<R$. The volume vanishes when $|x^5|=R$ at the north and south poles even though these are perfectly regular points of the four-geometry. One therefore would not expect the wave function to vanish at vanishing three-volume. Indeed, the three-volume will have to vanish somewhere if the topolo-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

gy of the four-geometry is not that of a product of a three-surface with the real line or the circle. When the volume does vanish, the topology of the three-geometry will change. One cannot calculate the amplitude for such topology change from the Wheeler-DeWitt equation but one can do so using the Euclidean functional integral. We shall estimate the amplitude in some simple cases in Sec. VIII.

A qualitative discussion of the expected behavior of the wave function at large three-volumes can be given on the basis of the semiclassical approximation when $\Lambda > 0$ as follows. The four-sphere has the largest volume of any real solution to (4.2). As the volume of the three-geometry becomes large one will reach three-geometries which no longer fit anywhere in the four-sphere. We then expect that the stationary-phase geometries become complex. The ground-state wave function will be a real combination of two expressions like (4.1) evaluated at the complex-conjugate stationary-phase four-geometries. We thus expect the wave function to oscillate as the volume of the three-geometry becomes large. If it oscillates without being strongly damped this corresponds to a universe which expands without limit.

The above considerations are only qualitative but do suggest how the behavior of the ground-state wave function determines the boundary conditions for the Wheeler-DeWitt equation. In the following we shall make these considerations concrete in a minisuperspace model.

V. MINISUPERSPACE MODEL

It is particularly straightforward to construct minisuperspace models using the functional-integral approach to quantum gravity. One simply restricts the functional integral to the restricted degrees of freedom to be quantized. In this and the following sections, we shall illustrate the general discussion of those preceding with a particularly simple minisuperspace model. In it we restrict the cosmological constant to be positive and the four-geometries to be spatially homogeneous, isotropic, and closed so that they are characterized by a single scale factor. An explicit metric in a useful coordinate system is

$$ds^2=\sigma^2[-N^2(t)dt^2+a^2(t)d\Omega_3^2]\,, \tag{5.1}$$

where $N(t)$ is the lapse function and $\sigma^2=l^2/24\pi^2$. For the matter degrees of freedom, we take a single conformally invariant scalar field which, consistent with the geometry, is always spatially homogeneous, $\phi=\phi(t)$. The wave function is then a function of only two variables:

$$\Psi=\Psi(a,\phi),\quad \Phi=\Phi(K,\tilde{\phi})\,. \tag{5.2}$$

Models of this general structure have been considered previously by DeWitt,[12] Isham and Nelson,[13] and Blyth and Isham.[14]

To simplify the subsequent discussion we introduce the following definitions and rescalings of variables:

$$\phi=\frac{\tilde{\phi}}{a}=\frac{\chi}{(2\pi^2\sigma^2)^{1/2}a}\,, \tag{5.3}$$

$$\Lambda=3\lambda/\sigma^2,\quad H^2=|\lambda|\,. \tag{5.4}$$

The Lorentzian action keeping χ and a fixed on the boundaries is

$$S^a=\tfrac{1}{2}\int dt\left[\frac{N}{a}\right]\left[\left[\frac{a}{N}\frac{da}{dt}\right]^2+a^2-\lambda a^4+\left[\frac{a}{N}\frac{d\chi}{dt}\right]^2-\chi^2\right]. \tag{5.5}$$

From this action the momenta π_a and π_χ conjugate to a and χ can be constructed in the usual way. The Hamiltonian constraint then follows by varying the action with respect to the lapse function and expressing the result in terms of a, χ, and their conjugate momenta. One finds

$$\tfrac{1}{2}(-\pi_a{}^2-a^2+\lambda a^4+\pi_\chi{}^2+\chi^2)=0\,. \tag{5.6}$$

The Wheeler-DeWitt equation is the operator expression of this classical constraint. There is the usual operator-ordering problem in passing from classical to quantum relations but its particular resolution will not be central to our subsequent semiclassical considerations. A class wide enough to remind oneself that the issue exists can be encompassed by writing

$$\pi_a{}^2=-\frac{1}{a^p}\frac{\partial}{\partial a}\left[a^p\frac{\partial}{\partial a}\right], \tag{5.7}$$

although this is certainly not the most general form possible. In passing from the classical constraint to its quantum operator form there is also the possibility of a matter-energy renormalization. This will lead to an additive arbitrary constant in the equation. We thus write for the quantum version of Eq. (5.6)

$$\frac{1}{2}\left[\frac{1}{a^p}\frac{\partial}{\partial a}\left[a^p\frac{\partial}{\partial a}\right]-a^2+\lambda a^4-\frac{\partial^2}{\partial\chi^2}+\chi^2-2\epsilon_0\right]\times\Psi(a,\chi)=0\,. \tag{5.8}$$

A useful property stemming from the conformal invariance of the scalar field is that this equation separates. If we assume reasonable behavior for the function Ψ in the amplitude of the scalar field we can expand in harmonic-oscillator eigenstates

$$\Psi(a,\chi)=\sum_n c_n(a)u_n(\chi)\,, \tag{5.9}$$

where

$$\frac{1}{2}\left[-\frac{d^2}{d\chi^2}+\chi^2\right]u_n(\chi)=(n+\tfrac{1}{2})u_n(\chi)\,. \tag{5.10}$$

The consequent equation for the $c_n(a)$ is

$$\frac{1}{2}\left[-\frac{1}{a^p}\frac{d}{da}\left[a^p\frac{dc_n}{da}\right]+(a^2-\lambda a^4)c_n\right]=(n+\tfrac{1}{2}-\epsilon_0)c_n\,. \tag{5.11}$$

For small a this equation has solutions of the form

$$c_n\approx\text{constant},\quad c_n\approx a^{1-p} \tag{5.12}$$

[if p is an integer there may be a $\log(a)$ factor]. For large a the possible behaviors are

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$c_n \sim a^{-(p/2+1)} \exp(\pm \tfrac{1}{3} i H a^3) \,. \tag{5.13}$$

To construct the solution of Eq. (5.11) which corresponds to the ground state of the minisuperspace model we turn to our Euclidean functional-integral prescription. As applied to this minisuperspace model, the prescription of Sec. III for $\Psi_0(a_0,\chi_0)$ would be to sum $\exp(-I[g,\phi])$ over those Euclidean geometries and field configurations which are represented in the minisuperspace and which satisfy the ground-state boundary conditions. The geometrical sum would be over compact geometries of the form

$$ds^2 = \sigma^2[d\tau^2 + a^2(\tau) d\Omega_3^2] \tag{5.14}$$

for which $a(\tau)$ matches the prescribed value of a_0 on the hypersurface of interest. The prescription for the matter field would be to sum over homogeneous fields $\chi(\tau)$ which match the prescribed value χ_0 on the surface and which are regular on the compact geometry. Explicitly we could write

$$\Psi_0(a_0,\chi_0) = \int \delta a \, \delta\chi \exp(-I[a,\chi]) \,, \tag{5.15}$$

where, defining $d\eta = d\tau/a$, the action is

$$I = \tfrac{1}{2} \int d\eta \left[-\left| \frac{da}{d\eta} \right|^2 - a^2 + \lambda a^4 + \left| \frac{d\chi}{d\eta} \right|^2 + \chi^2 \right] . \tag{5.16}$$

A conformal rotation [in this case of $a(\eta)$] is necessary to make the functional integral in (5.15) converge.[15]

An alternative way of constructing the ground-state wave function for the minisuperspace model is to work in the K representation. Here, introducing

$$k = \sigma K / 9 \tag{5.17}$$

as a simplifying measure of K, one would have

$$\Phi_0(k_0,\chi_0) = \int \delta a \, \delta\chi \exp(-I^k[a,\chi]) \,. \tag{5.18}$$

The sum is over the same class of geometries and fields as in (5.15) except they must now assume the given value of k on the bounding three-surface. That is, on the boundary they must satisfy

$$k_0 = \frac{1}{3a} \frac{da}{d\tau} \,. \tag{5.19}$$

The action I^k appropriate for holding k fixed on the boundary is

$$I^k = k_0 a_0^3 + I \tag{5.20}$$

[cf. Eq. (3.6)]. Once $\Phi_0(k_0,\chi_0)$ has been computed, the ground-state wave function $\Psi_0(k_0,\chi_0)$ may be recovered by carrying out the contour integral

$$\Psi_0(a_0,\chi_0) = -\frac{1}{2\pi i} \int_C dk \, e^{k a_0^3} \Phi_0(k,\chi_0) \,, \tag{5.21}$$

where the contour runs from $-i\infty$ to $+i\infty$ to the right of any singularities of $\Phi_0(k_0,\chi_0)$.

From the general point of view there is no difference between computing $\Psi_0(a_0,\chi_0)$ directly from (5.15) or via the K representation from (5.21). In Sec. VI we shall calculate the semiclassical approximation to $\Psi_0(a_0,\chi_0)$ both ways with the aim of advancing arguments that the rules of Sec. III define a wave function which may reasonably be considered as the state of minimal excitation and of displaying the boundary conditions under which Eq. (5.11) is to be solved.

VI. GROUND-STATE COSMOLOGICAL WAVE FUNCTION

In this section, we shall evaluate the ground-state wave function for our minisuperspace model and show that it possesses properties appropriate to a state of minimum excitation. We shall first evaluate the wave function in the semiclassical approximation from the steepest-descents approximation to the defining functional integral as described in Sec. IV. We shall then solve the Wheeler-DeWitt equation with the boundary conditions implied by the semiclassical approximation to obtain the precise wave function.

It is the exponent of the semiclassical approximation which will be most important in its interpretation. We shall calculate only this exponent from the extrema of the action and leave the determination of the prefactor [cf. Eq. (4.1)] to the solution of the differential equation. Thus, for example, if there were a single real Euclidean extremum of least action we would write for the semiclassical approximation to the functional integral in Eq. (5.15)

$$\Psi_0(a_0,\chi_0) \approx N e^{-I(a_0,\chi_0)} \,. \tag{6.1}$$

Here, $I(a_0,\chi_0)$ is the action (5.16) evaluated at the extremum configurations $a(\tau)$ and $\chi(\tau)$ which satisfy the ground-state boundary conditions spelled out in Sec. III and which match the arguments of the wave function on a fixed-τ hypersurface.

A. The matter wave function

A considerable simplification in evaluating the ground-state wave function arises from the fact that the energy-momentum tensor of an extremizing conformally invariant field vanishes in the compact geometries summed over as a consequence of the ground-state boundary conditions. One can see this because the compact four-geometries of the class we are considering are conformal to the interior of three-spheres in flat Euclidean space. A constant scalar field is the only solution of the conformally invariant wave equation on flat space which is a constant on the boundary three-sphere. The energy-momentum tensor of this field is zero. This implies that it is zero in any geometry of the class (5.14) because the energy-momentum tensor of a conformally invariant field scales by a power of the conformal factor under a conformal transformation.

More explicitly in the minisuperspace model we can show that the matter and gravitational functional integrals in (5.15) may be evaluated separately. The ground-state boundary conditions imply that geometries in the sum are conformal to half of a Euclidean Einstein-static universe, i.e., that the range of η is $(-\infty,0)$. The boundary conditions at infinite η are that $\chi(\eta)$ and $a(\eta)$ vanish. The boundary conditions at $\eta=0$ are that $a(0)$ and $\chi(0)$ match

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the arguments of the wave function a_0 and χ_0. Thus, not only does the action (5.16) separate into a sum of a gravitational part and a matter part, but the boundary conditions on the $a(\eta)$ and $\chi(\eta)$ summed over do not depend on one another. The matter and gravitational integrals can thus be evaluated separately.

Let us consider the matter integral first. In Eq. (5.16) the matter action is

$$I_M=\tfrac{1}{2}\int d\eta\left[\left(\frac{d\chi}{d\eta}\right)^2+\chi^2\right]. \tag{6.2}$$

This is the Euclidean action for the harmonic oscillator. Evaluation of the matter field integral in (5.15) therefore gives

$$\Psi_0(a_0,\chi_0)=e^{-\chi_0^2/2}\psi_0(a_0)\,. \tag{6.3}$$

Here, $\psi_0(a)$ is the wave function for gravity alone given by

$$\psi_0(a_0)=\int \delta a\,\exp(-I_E[a])\,, \tag{6.4}$$

I_E being the gravitational part of (5.16). Equivalently we can write in the K representation

$$\Phi_0(k_0,\chi_0)=e^{-\chi_0^2/2}\phi_0(k_0)\,, \tag{6.5}$$

where

$$\phi_0(k_0)=\int \delta a\,\exp(-I_E^k[a])\,. \tag{6.6}$$

$I_E{}^k[a]$ is related to I_E as in (5.20) and the sum is over $a(\tau)$ which satisfy (5.19) on the boundary. Equation (6.3) shows that as far as the matter field is concerned, $\Psi_0[a_0,\chi_0]$ is reasonably interpreted as the ground-state wave function. The field oscillators are in their state of minimum excitation—the ground state of the harmonic oscillator. We now turn to a semiclassical calculation of the gravitational wave function $\psi_0(a_0)$.

B. The semiclassical ground-state gravitational wave function

The integral in (6.4) is over $a(\tau)$ which represent [through (5.14)] compact geometries with three-sphere boundaries of radius a. The integral in (6.6) is over the same class of geometries except that the three-sphere boundary must possess the given value of k. The compact geometry which extremizes the gravitational action in these cases is a part of the Euclidean four-sphere of radius $1/H$ with an appropriate three-sphere boundary. In the case where the three-sphere radius is fixed on the boundary there are two extremizing geometries. For one the part of the four-sphere bounded by the three-sphere is greater than a hemisphere and for the other it is less. A careful analysis must therefore be made of the functional integral to see which of these extrema contributes to the semiclassical approximation. We shall give such an analysis below but first we show that the correct answer is achieved more directly in the K representation from (6.6) because there is a single extremizing geometry with a prescribed value of k on a three-sphere boundary and thus no ambiguity in constructing the semiclassical approximation to (6.6).

For three-sphere hypersurfaces of the four-sphere with an outward pointing normal, k ranges from approaching $+\infty$ for a surface encompassing a small region about a pole to approaching $-\infty$ for the whole four-sphere (see Fig. 1). More exactly, in the notation of Eq. (3.7)

$$k=\frac{H}{3}\cot\theta\,. \tag{6.7}$$

The extremum action is constructed through (5.20) with the integral in (5.16) being taken over that part of the four-sphere bounded by the three-sphere of given k. It is

$$I_E^k(k)=-\frac{1}{3H^2}\left[1-\frac{\kappa}{(\kappa^2+1)^{1/2}}\right], \tag{6.8}$$

where

$$k=\tfrac{1}{3}\kappa H\,. \tag{6.9}$$

The semiclassical approximation to (6.6) is now

$$\phi_0(k_0)\approx N\exp[-I_E^k(k_0)]\,. \tag{6.10}$$

The wave function $\psi_0(a_0)$ in the same approximation can be constructed by carrying out the contour integral

$$\psi_0(a_0)=-\frac{N}{2\pi i}\int_C dk\,\exp[ka_0{}^3-I_E{}^k(k)] \tag{6.11}$$

by the method of steepest descents. The exponent in the integrand of Eq. (6.11) is minus the Euclidean action for pure gravity with a kept fixed instead of k:

$$I_E(a)=-ka^2+I_E{}^k(k)\,. \tag{6.12}$$

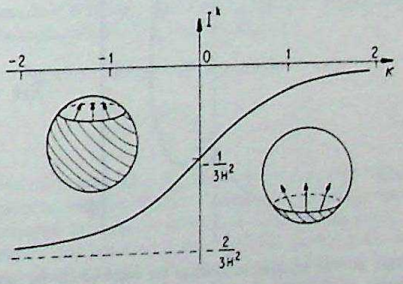


FIG. 1. The action I^k for the Euclidean four-sphere of radius $1/H$. The Euclidean gravitational action for the part of a four-sphere bounded by a three-sphere of definite K is plotted here as a function of κ (a dimensionless measure of K [Eq. (6.9)]). The action is that appropriate for holding K fixed on the boundary. The shaded regions of the inset figures show schematically the part of the four-sphere which fills in the three-sphere of given K used in computing the action. A three-sphere of given K fits in a four-sphere at only one place. Three-spheres with positive K (diverging normals) bound less than a hemisphere of four-sphere while those with negative K (converging normals) bound more than a hemisphere. The action tends to its flat-space value (zero) as K tends to positive infinity. It tends to the Euclidean action for all of de Sitter space as K tends to negative infinity.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

To evaluate (6.11) by steepest descents we must find the extrema of Eq. (6.12). There are two cases depending on whether Ha_0 is greater or less than unity.

For $Ha_0 < 1$ the extrema of $I_E(k)$ occur at real values of k which are equal in magnitude and opposite in sign. They are the values of k at which a three-sphere of radius a_0 would fit into the four-sphere of radius $1/H$. That is, they are those values of k for which Eq. (6.7) is satisfied with $a_0^2 = (\sin\theta/H)^2$. This is not an accident; it is a consequence of the Hamilton-Jacobi theory. The value of I_E at these extrema is

$$I_\pm = -\frac{1}{3H^2}[1 \pm (1-H^2a_0^2)^{3/2}] , \tag{6.13}$$

where the upper sign corresponds to $k<0$ and the lower to $k>0$, i.e., to filling in the three-sphere with greater than a hemisphere of the four-sphere or less than a hemisphere, respectively.

There are complex extrema of I_E but all have actions whose real part is greater than the real extrema described above. The steepest-descents approximation to the integral (6.11) is therefore obtained by distorting the contour into a steepest-descents path (or sequence of them) passing through one or the other of the real extrema. The two real extrema and the corresponding steepest-descents directions are shown in Fig. 2. One can distort the contour into a steepest-descents path passing through only one of them—the one with positive k as shown. The functional integral thus singles out a unique semiclassical approximation to $\psi_0(a_0)$ which is

$$\psi_0(a_0) \approx N \exp[-I_-(a_0)], \quad Ha_0 < 1 , \tag{6.14}$$

corresponding to filling in the three-sphere with less than a hemisphere's worth of four-sphere.

From Eq. (4.4) we recover the normalization factor N:

$$N = \exp(-\tfrac{1}{3}H^{-2}) . \tag{6.15}$$

Thus, for $Ha_0 \ll 1$

$$\psi_0(a_0) = \exp(\tfrac{1}{2}a_0^2 - \tfrac{1}{3}H^{-2}) . \tag{6.16}$$

One might have thought that the extremum I_+, which corresponds to filling in the three-geometry with more than a hemisphere, would provide the dominant contribution to the ground-state wave function as $\exp(-I_+)$ is greater than $\exp(-I_-)$. However, the steepest-descents contour in the integral (6.7) does not pass through the extremum coresponding to I_+. This is related to the fact that the contour of integration of the conformal factor has to be rotated in the complex plane in order to make the path integral converge as we shall show below.

For $Ha_0 > 1$ there are no real extrema because we cannot fit a three-sphere of radius $a_0 > 1/H$ into a four-sphere of radius $1/H$. There are, however, complex extrema of smallest real action located at

$$k = \pm\frac{i}{3}H\left[1 - \frac{1}{H^2a_0^2}\right]^{1/2} . \tag{6.17}$$

It is possible to distort the contour in Eq. (6.11) into a steepest-descents contour passing through both of them as shown in Fig. 3. The resulting wave function has the form

$$\psi_0(a_0) = 2\cos\left[\frac{(H^2a_0^2-1)^{3/2}}{3H^2} - \frac{\pi}{4}\right] , \quad Ha_0 > 1 \tag{6.18}$$

or for $Ha_0 \gg 1$

$$\psi_0(a_0) \approx e^{+iHa_0^3/3} + e^{-iHa_0^3/3} . \tag{6.19}$$

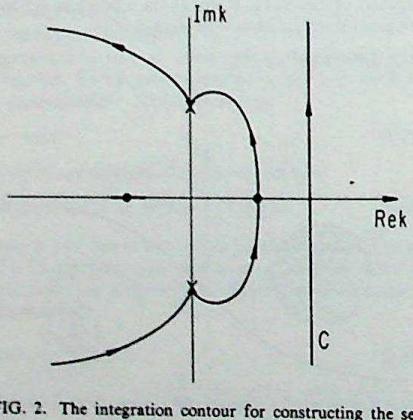


FIG. 2. The integration contour for constructing the semiclassical ground-state wave function of the minisuperspace model in the case $\Lambda > 0$, $Ha_0 < 1$. The figure shows schematically the original integration contour C used in Eq. (6.11) and the steepest-descents contour into which it can be distorted. The branch points of the exponent of Eq. (6.11) at $\kappa = \pm i$ are located by crosses. There are two extrema of the exponent which correspond to filling in the three-sphere of given radius a with greater than a hemisphere of four-sphere or less than a hemisphere. For $Ha_0 < 1$ they lie at the equal and opposite real values of K indicated by dots. The contour C can be distorted into a steepest-descents contour through the extremum with positive K as shown. It cannot be distorted to pass through the extremum with negative K in the steepest-descents direction indicated. The contour integral thus picks out the extremum corresponding to less than a hemisphere of four-sphere (cf. Fig. 1) as the leading term in the semiclassical approximation.

The semiclassical approximation to the ground-state gravitational wave function $\psi_0(a)$ contained in Eqs. (6.16) and (6.19) may also be obtained directly from the functional integral (6.4) without passing through the k representation. We shall now sketch this derivation. We must consider explicitly the conformal rotation which makes the gravitational part of the action in (5.16) positive definite. The gravitational action is

$$I_E[a] = \tfrac{1}{2}\int d\eta \left[-\left[\frac{da}{d\eta}\right]^2 - a^2 + H^2a^4\right] . \tag{6.20}$$

If one performed the functional integration

$$\psi_0(a_0) = \int \delta a(\eta) \exp(-I_E[a]) \tag{6.21}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

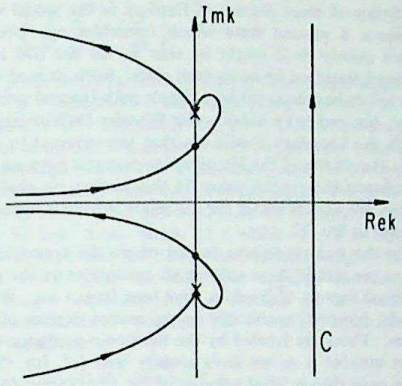


FIG. 3. The integration contour for constructing the semiclassical ground-state wave function of the minisuperspace model in the case $\Lambda > 0$, $Ha_0 > 1$. The figure shows schematically the original contour C used in Eq. (6.11) and the steepest-descents contour into which it can be distorted. The branch points of the exponent of Eq. (6.11) at $\kappa = \pm i$ are located by crosses. There are two complex-conjugate extrema of the exponent as indicated by dots and the contour C can be distorted to pass through both along the steepest-descents directions at 45° to the real axis as shown.

over real values of a, one would obtain a divergent result because the first term in (6.20) is negative definite. One could make the action infinitely negative by choosing a rapidly varying a. The solution to this problem seems to be to integrate the variable a in Eq. (6.21) along a contour that is parallel to the imaginary axis.[15] For each value of η, the contour of integration of a will cross the real axis at some value. Suppose there is some real function $\bar{a}(\eta)$ which maximizes the action. Then if one distorts the contour of integration of a at each value of η so that it crosses the real axis at $\bar{a}(\eta)$, the value of the action at the solution $\bar{a}(\eta)$ will give the saddle-point approximation to the functional integral (6.21), i.e.,

$$\psi_0(a_0) \approx N \exp(-I_E[\bar{a}(\eta)]) . \tag{6.22}$$

If there were another real function $\hat{a}(\eta)$ which extremized the action but which did not give its maximum value there would be a nearby real function $\hat{a}(\eta)+\delta a(\eta)$ which has a greater action. By choosing the contour of integration in (6.21) to cross the real a axis at $\hat{a}(\eta)+\delta a(\eta)$, one would get a smaller contribution to the ground-state wave function. Thus, the dominant contribution comes from the real function $\bar{a}(\eta)$ with the greatest value of the action.

It may be that there is no real $a(\eta)$ which maximizes the action. In this case the dominant contribution to the ground-state wave function will come from complex functions $a(\eta)$ which extremize the action. These will occur in complex-conjugate pairs because the wave function is real.

In the case of $Ha_0 < 1$, we have already seen that there are two real functions $a(\eta)$ which extremize the action and which correspond to less than or more than a hemisphere of the four-sphere. Their actions are I_- and I_+, respectively, given by (6.13). In fact, I_- is the maximum value of the action for real $a(\eta)$ and therefore gives the dominant contribution to the ground-state wave function. Thus, we again recover Eqs. (6.14) and (6.16). In the case of $Ha_0 > 1$, there is no maximum of the action for real $a(\eta)$. In this case the dominant contribution to the ground-state wave function comes from a pair of complex-conjugate $a(\eta)$ which extremize the action. Thus, we would expect an oscillatory wave function like that given by Eq. (6.19).

C. Ground-state solution of the Wheeler-DeWitt equation

The ground-state wave function must be a solution of the Wheeler-DeWitt equation for the minisuperspace model [Eqs. (5.8) or (5.11)]. The $\exp(-\chi^2/2)$ dependence of the wave function on the matter field deduced in Sec. VI A shows that in fact $\psi_0(a)$ must solve Eq. (5.11) with $n=0$. There are certainly solutions of this equation which have the large-a combination of exponentials required of the semiclassical approximation by Eq. (6.19) as a glance at Eq. (5.13) shows. In fact the prefactor in these asymptotic behaviors shows that the ground-state wave function will be *normalizable* in the norm

$$(\psi_0', \psi_0) = \int_0^\infty da\, a^p \bar{\psi}_0'(a) \psi_0(a) \tag{6.23}$$

in which the Wheeler-DeWitt operator is Hermitian.

The Wheeler-DeWitt equation enables us to determine the prefactor in the semiclassical approximation from the standard WKB-approximation formulas. With $p=0$, for example, this would give when $Ha_0 > 1$

$$\psi_0(a_0) = 2(H^2 a_0^4 - a_0^2 + \epsilon_0 + \tfrac{1}{2})^{-1/4} \times \cos\left[\frac{(H^2 a_0^2 - 1)^{3/2}}{3H^2} - \frac{\pi}{4} \right] . \tag{6.24}$$

We could also solve the equation numerically. Figure 4 gives an example when $p=0$ and $\epsilon_0 = -\frac{1}{2}$. There we have assumed that the wave function vanishes at $a=0$. The dotted lines represent graphs of the prefactor in Eq. (6.24) and show that the semiclassical approximation becomes rapidly more accurate as Ha increases beyond 1. We shall return to an interpretation of these facts below.

D. Correspondence with de Sitter space

Having obtained $\psi_0(a)$, we are now in a position to assess its suitability as the ground-state wave function. Classically the vacuum geometry with the highest symmetry, hence minimum excitation, is de Sitter space—the surface of a Lorentz hyperboloid in a five-dimensional Lorentz-signatured flat spacetime. The properties of the wave function contained in Eqs. (6.16) and (6.19) are those one would expect to be semiclassically associated with this geometry. Sliced into three-spheres de Sitter space contains spheres only with a radius greater than $1/H$. Equation (6.16) shows that the wave function is an exponential-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

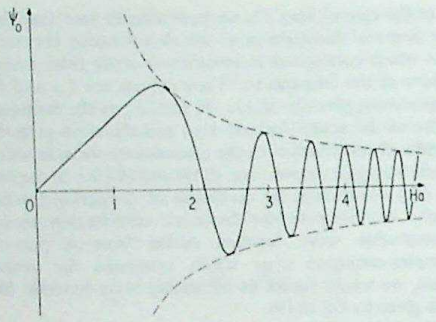


FIG. 4. A numerical solution of the Wheeler-DeWitt equation for the ground-state wave function $\psi_0(a)$. A solution of Eq. (5.11) is shown for $H=1$ in Planck units. We have assumed for definiteness $p=0$, $\epsilon_0=-\frac{1}{2}$, and a vanishing wave function at the origin. The wave function is damped for $Ha<1$ corresponding to the absence of spheres of radii smaller than H^{-1} in Lorentzian de Sitter space. It oscillates for $Ha>1$ decaying only slowly for large a. This reflects the fact that de Sitter space expands without limit. In fact, the envelope represented by the dotted lines is the distribution of three-spheres in Lorentzian de Sitter space: $[Ha(H^2a^2-1)^{1/2}]^{-1}$.

ly decreasing function with decreasing a for radii below that radius. Equation (6.24) shows the spheres of radius larger than $1/H$ are found with an amplitude which varies only slowly with the radius. This is a property expected of de Sitter space which expands both to the past and the future without limit. Indeed, tracing the origin of the two terms in (6.19) back to extrema with different signs of k one sees that one of these terms corresponds to the contracting phase of de Sitter space while the other corresponds to the expanding phase. The slow variation in the amplitude of the ground-state wave function reflects precisely the distribution of three-spheres in classical de Sitter space. Lorentzian de Sitter space is conformal to a finite region of the Einstein static universe

$$ds^2=\sigma^2a^2(\eta)(-d\eta^2+d\Omega_3^2)\,, \tag{6.25}$$

where $a(t)=(\cosh Ht)/H$ and $dt=ad\eta$. Three-spheres are evidently distributed uniformly in η in the Einstein static universe. The distribution of spheres in a in Lorentzian de Sitter space is therefore proportional to

$$[a(H^2a^2-1)^{1/2}]^{-1}\,. \tag{6.26}$$

This is the envelope of the probability distribution $a^p|\psi(a)|^2$ for spheres of radius a deduced from the semiclassical wave function and shown in Fig. 4. The wave function constructed from the Euclidean prescription of Sec. III appropriately reflects the properties of the classical vacuum solution of highest symmetry and is therefore reasonably called the ground-state wave function.

VII. EXCITED STATES

Our Universe does not correspond to the ground state of the simple minisuperspace model. It might be that the inclusion of more degrees of freedom in the model would produce a ground state which resembles our Universe more closely or it might be that we do not live in the ground state but in an excited state. Such excited states are not to be calculated by a simple path-integral prescription, but rather by solving the Wheeler-DeWitt equation with the boundary conditions that are required to maintain Hemiticity of the Hamiltonian operator between these states and the ground state. In this section, we shall construct the excited states for the minisuperspace model discussed in Sec. VI.

In the minisuperspace model where the spacelike sections are metric three-spheres all excitations in the gravitational degrees of freedom have been frozen out. We can study, however, excitations in the matter degrees of freedom. These are labeled by the harmonic-oscillator quantum number n as we have already seen [cf. Eq. (5.10)]. The issue then is what solution of Eq. (5.11) for $c_n(a)$ corresponds to this excited state. The equation can be written in the form of a one-dimensional Schrödinger equation

$$-\frac{1}{2a^p}\frac{d}{da}\left[a^p\frac{dc_n}{da}\right]+V(a)c_n=(n+\tfrac{1}{2}+\epsilon_0)c_n\,, \tag{7.1}$$

where

$$V(a)=\tfrac{1}{2}(a^2-\lambda a^4)\,. \tag{7.2}$$

At $a=0$ Eq. (7.1) will in general have two types of solutions one of which is more convergent than the other [cf. Eq. (5.12)]. The behavior for the ground state which corresponds to the functional-integral prescription could be deduced from an evaluation of the determinant in the semiclassical approximation as discussed in Sec. IV. Whatever the result of such an evalution, the solution must be purely of one type or the other in order to ensure the Hermiticity of the Hamiltonian constraint. The same requirement ensures a similar behavior for the excited-state solutions. In the following by "regular" solutions we shall mean those conforming to the boundary conditions arising from the functional-integral prescription. The exact type will be unimportant to us.

The potential $V(a)$ is a barrier of height $1/(4\lambda)$. At large a, the cosmological-constant part of the potential dominates and one has solutions which are linear combinations of the oscillating functions in (5.13). As we have already seen in the analysis of the ground state, the two possibilities correspond to a de Sitter contraction and a de Sitter expansion. With either of these asymptotic behaviors, a wave packet constructed by superimposing states of different n to produce a wave function with narrow support about some mean value of the scalar field would show this mean value increasing as one moved from large to small a.

Since each of the asymptotic behaviors in (5.13) is physically acceptable there will be solutions of (7.1) for all n. If, however, λ is small and n not too large, there are some values of n which are more important than others. These are the values which make the left-hand side of (7.1) at or close to those values of the energy associated with the metastable states (resonances) of the Schrödinger Hamiltonian on the right-hand side. To make this precise write

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$-\frac{1}{2}\frac{1}{a^p}\frac{d}{da}\left[a^p\frac{dc}{da}\right]+V(a)c=\epsilon c\ . \tag{7.3}$$

This is the zero angular momentum Schrödinger equation in $d=p+1$ dimensions for single-particle motion in the potential $V(a)$. Classically, for $\epsilon<1/(4\lambda)$ there are two classes of orbits: bound orbits with a maximum value of a and unbound orbits with a minimum value of a. Quantum mechanically there are no bound states. For discrete values of $\epsilon\ll 1/(4\lambda)$, however, there are metastable states. They lie near those values of ϵ which would be bound states if $\lambda=0$ and the barrier had infinite height. Since when $\lambda=0$ (7.3) is the zero angular momentum Schrödinger equation for a particle in a "radial" harmonic-oscillator potential in $d=p+1$ dimensions, these values are

$$\epsilon_N=2N+d/2,\quad N=0,1,2,\ldots\ . \tag{7.4}$$

For nonzero λ, if the particle has an energy near one of these values and much less than $1/(4\lambda)$ it can execute many oscillations inside the well but eventually it will tunnel out.

For the cosmological problem the classical Hamiltonian corresponding to (7.3) describes the evolution of homogeneous, isotropic, spatially closed cosmologies with radiation and a cosmological constant. The bound orbits correspond to those solutions for which the radiation density is sufficiently high that its attractive effect causes an expanding universe to recollapse before the repulsive effect of the cosmological constant becomes important. By contrast the unbound orbits correspond to de Sitter evolutions in which a collapsing universe never reaches a small enough volume for the increasing density of radiation to reverse the effect of the cosmological constant. There are thus two possible types of classical solutions. Quantum mechanically the Universe can tunnel between the two.

We can calculate the tunneling probability for small λ by using the usual barrier-penetration formulas from ordinary quantum mechanics. Let P be the probability for tunneling from inside the barrier to outside per transversal of the potential inside from minimum to maximum a. Then

$$P\approx e^{-B}\ , \tag{7.5}$$

where

$$B=2\int_{a_0}^{a_1}da[V(a)-\epsilon]^{1/2} \tag{7.6}$$

and a_0 and a_1 are the two turning points where $V(a)=\epsilon$. In the limit of $\epsilon\ll 1/(4\lambda)$ the barrier-penetration factor becomes

$$B=\frac{2}{3\lambda}=\frac{2}{3H^2}\ . \tag{7.7}$$

In magnitude this is just the total gravitational action for the Euclidean four-sphere of radius $1/H$ which is the analytic continuation of de Sitter space. This is familiar from general semiclassical results.[16]

Our own Universe corresponds to a highly excited state of the minisuperspace model. We know that the age of the Universe is about 10^{60} Planck times. The maximum expansion, assuming a radiation dominated model, is therefore at least of order $a_{\max}{}^2\approx 10^{120}$. A wave packet describing our Universe would therefore have to be superpositions of states of definite n, with n at least $\approx a_{\max}{}^2\approx 10^{120}$. As large as this number is, the dimensionless limit on the inverse cosmological constant is even larger. In order to have such a large radiation dominated Universe λ must be less than 10^{-120}. The probability for our Universe to tunnel quantum mechanically at the moment of its maximum expansion to a de Sitter-type phase rather than recollapse is $P\approx\exp(-10^{120})$. This is a very small number but of interest if only because it is nonzero.

VIII. TOPOLOGY

In the preceding sections we have considered the amplitudes for three-geometries with S^3 topology to occur in the ground state. The functional-integral construction of the ground-state wave function, however, permits a natural extension to calculate the amplitudes for other topologies. We shall illustrate this extension in this section with some simple examples in the semiclassical approximation.

There is no compelling reason for restricting the topologies of the Euclidean four-geometries which enter in the sum defining the ground-state wave function. Whatever one's view on this question, however, there must be a ground-state wave function for every topology of a three-geometry which can be embedded in a four-geometry which enters the sum. In the general case this will mean all possible three-topologies—disconnected as well as connected, multiply connected as well as simply connected. The general ground-state wave function will therefore have N arguments representing the possibility of N compact disconnected three-geometries. The functional-integral prescription for the ground-state wave function in the case of pure gravity would then read

$$\Psi_0[\partial M^{(1)},h_{ij}^{(1)},\ldots,\partial M^{(N)},h_{ij}^{(N)}]=\int\delta g\exp(-I_E[g])\ , \tag{8.1}$$

where the sum is over all compact Euclidean four-geometries which have N disconnected compact boundaries $\partial M^{(i)}$ on which the induced three-metrics are $h_{ij}^{(i)}$. Since there is nothing in the sum which distinguishes one three-boundary from another the wave function must be symmetric in its arguments.

The wave function defined by (8.1) obeys a type of Wheeler-DeWitt equation in each argument but this is no longer sufficient to determine its form—in particular the correlations between the three-geometries. The functional integral is here the primary computational tool.

It is particularly simple to construct the semiclassical approximations to ground-state wave functions for those three-geometries with topologies which can be embedded in a compact Euclidean solution of the field equations. Consider for example the four-sphere. If the three-geometry has a single connected component and can be embedded in the four-sphere, then the extremal geometry at which the action is evaluated to give the semiclassical approximation is the smaller part of the four-sphere bounded by this three-geometry. The semiclassical ground-state wave function is

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$\Psi_0[h_{ij}] \approx N\Delta^{-1/2}[h_{ij}] \times \exp\left[2l^{-2}\left(\int_{\partial M} dx^3 h^{1/2}K + \int_M d^4x\, g^{1/2}\Lambda\right)\right] \tag{8.2}$$

where M is the smaller part of the four-sphere and K is the trace of the extrinsic curvature of the three-surface computed with outward-pointing normals. Since there is a large variety of topologies of three-surfaces which *can* be embedded in the four-sphere—spheres, toruses, etc.,—we can easily compute their associated wave functions. Of course, these are many interesting three-surfaces which cannot be so embedded and for which the extremal solution defining the semiclassical approximation is not part of the four-sphere. In general one would expect to find wave functions for arbitrary topologies since any three-geometry is cobordant to zero and therefore there is some compact four-manifold which has it as its boundary. The problem of finding solutions of the field equations on these four-manifolds which match the given three-geometry and are compact thus becomes an interesting one.

Similarly, the semiclassical approximation for wave functions representing N disconnected three-geometries are equally easily computed when the geometries can be embedded in the four-sphere. The extremal geometry defining the semiclassical approximation is then simply the four-sphere with the N three-geometries cut out of it. The symmetries of the solution guarantee that as far as the exponent of the semiclassical approximation is concerned, it does not matter where the three-geometries are cut out provided that they do not overlap. To give a specific example, we calculate the amplitude for two disconnected three-spheres of radius $a_{(1)}$ and $a_{(2)}$ assuming $a_{(1)} < a_{(2)} < H^{-1}$. One possible extremal geometry is two disconnected portions of a four-sphere attached to the two three-spheres. This gives a product wave function with no correlation. Another extremal geometry is the smaller half of the four-sphere bounded by the spheres of radius $a_{(2)}$ with the portion interior to a sphere of radius $a_{(1)}$ removed. This gives an additional contribution to the wave function which expresses the correlation between the spheres. The correlated part in the semiclassical approximation is

$$\Psi_0^c(a_{(1)},a_{(2)}) = N\Delta^{-1/2}(a_{(1)},a_{(2)}) \times \exp\left\{\frac{1}{3H^2}[-(1-H^2a_{(2)}^2)^{3/2} + (1-H^2a_{(1)}^2)^{3/2}]\right\}. \tag{8.3}$$

While the exponent is simple, the calculation of the determinant is now more complicated—it does not factor.

Equation (8.3) shows that the amplitude to have two correlated three-spheres of radius $a_{(1)} < a_{(2)} < H^{-1}$ is smaller than the amplitude to have a single three-sphere of radius $a_{(2)}$. In this crude sense topological complexity is suppressed. The amplitude for the Universe to bifurcate is of the order $\exp[-1/(3H^2)]$—a very large factor.

IX. CONCLUSIONS

The ground-state wave function for closed universes constructed by the Euclidean functional-integral prescription put forward in this paper can be said to represent a state of minimal excitation for these universes for two reasons. First, it is the natural generalization to gravity of the Euclidean functional integral for the ground-state wave function of flat-spacetime field theories. Second, when the prescription is applied to simple minisuperspace models, it yields a semiclassical wave function which corresponds to the classical solution of Einstein's equations of highest spacetime symmetry and lowest matter excitation.

The advantages of the Euclidean function-integral prescription are many but perhaps three may be singled out. First it is a complete prescription for the wave function. It implies not only the Wheeler-DeWitt equation but also the boundary conditions which determine the ground-state solution. The requirement of Hermiticity of the Wheeler-DeWitt operator extends these boundary conditions to the excited states as well.

A second advantage of this prescription for the ground-state wave function is common to all functional-integral formulations of quantum amplitudes. They permit the direct and explicit calculation of the semiclassical approximation. At the current stage of the development of quantum gravity where qualitative understanding is more important than precise numerical results, this is an important advantage. It is well illustrated by our minisuperspace model in which we were able to calculate semiclassically the probability of tunneling between a universe doomed to end in a big crunch and an eternal de Sitter expansion.

A final advantage of the Euclidean functional-integral prescription for the ground-state wave function is that it naturally generalizes to permit the calculation of amplitudes not usually considered in the canonical theory. In particular, we have been able to provide a functional-integral prescription for amplitudes for the occurrence of three-geometries with multiply connected and disconnected topologies in the ground state. In the semiclassical approximation we have been able to evaluate simple examples of such amplitudes.

The Euclidean functional-integral prescription sheds light on one of the fundamental problems of cosmology: the singularity. In the classical theory the singularity is a place where the field equations, and hence predictability, break down. The situation is improved in the quantum theory. An analogous improvement occurs in the problem of an electron orbiting a proton. In the classical theory there is a singularity and a breakdown of predictability when the electron is at the same position as the proton. However, in the quantum theory there is no singularity or breakdown. In an s-wave state, the amplitude for the electron to coincide with the proton is finite and nonzero, but the electron just carries on to the other side. Similarly, the amplitude for a zero-volume three-sphere in our minisuperspace model is finite and nonzero. One might interpret this as implying that the universe could continue through the singularity to another expansion period, although the classical concept of time would break down so that one

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

could not say that the expansion happened after the contraction.

The ground-state wave function in the simple minisuperspace model that we have considered with a conformally invariant field does not correspond to the quantum state of the Universe that we live in because the matter wave function does not oscillate. However, it seems that this may be a consequence of using only zero rest mass fields and that the ground-state wave function for a universe with a massive scalar field would be much more complicated and might provide a model of quantum state of the observed Universe. If this were the case, one would have solved the problem of the initial boundary conditions of the Universe: the boundary conditions are that it has no boundary.[3]

ACKNOWLEDGMENTS

The authors are grateful for the hospitality of the Institute of Theoretical Physics, Santa Barbara, California, where part of this work was carried out. The research of one of us (J.H.) was supported in part by NSF Grants Nos. PHY81-07384 and PHY80-26043.

---

[1]See, e.g., R. P. Feynman, Rev. Mod. Phys. 20, 367 (1948); R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965) for discussions of quantum mechanics from this point of view.

[2]B. S. DeWitt, Phys. Rev. 160, 1113 (1967); J. A. Wheeler, in *Battelle Rencontres,* edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

[3]S. W. Hawking, in *Astrophysical Cosmology,* Pontificia Academiae Scientarium Scripta Varia, 48' (Pontificia Academiae Scientarium, Vatican City, 1982).

[4]For related ideas, see A. Vilenkin, Phys. Lett. 117B, 25 (1982); Phys. Rev. D 27, 2848 (1983).

[5]The connection between the canonical and functional-integral approaches to quantum gravity has been extensively discussed. See, in particular, H. Leutwyler, Phys. Rev. 134, B1155 (1964); L. Faddeev and V. Popov, Usp. Fiz. Nauk. 111, 427 (1973) [Sov. Phys. Usp. 16, 777 (1974)]; E. S. Fradkin and G. A. Vilkovisky, CERN Report No. TH-2332, 1977 (unpublished).

[6]For reviews of the canonical theory, see K. Kuchar, in *Quantum Gravity 2,* edited by C. Isham, R. Penrose, and D. W. Sciam (Clarendon, Oxford, 1981); A Hanson, T. Regge, and C. Teiteiboim, *Constrained Hamiltonian Systems* (Academia Nazionale dei Lincei, Rome, 1976); K. Kuchar, in *Relativity, Astrophysics and Cosmology,* edited by W. Israel (Reidel, Dordrecht, 1973).

[7]J. M. York, Phys. Rev. Lett. 28, 1082 (1972).

[8]J. Milnor, *Morse Theory* (Princeton University Press, Princeton, New Jersey, 1962).

[9]For example, J. B. Hartle (unpublished).

[10]See, e.g., G. W. Gibbons and M. J. Perry, Nucl. Phys. B146, 90 (1978).

[11]S. W. Hawking, Commun. Math. Phys. 55, 133 (1977).

[12]B. DeWitt, Phys. Rev. 160, 1113 (1967).

[13]C. Isham and J. E. Nelson, Phys. Rev. D 10, 3226 (1974).

[14]W. E. Blyth and C. Isham, Phys. Rev. D 11, 768 (1975).

[15]G. W. Gibbons, S. W. Hawking, and M. J. Perry, Nucl. Phys. B138, 141 (1978).

[16]See, e.g., S. Coleman, Phys. Rev. D 15, 2929 (1977).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

COURSE 4

QUANTUM COSMOLOGY

S.W. HAWKING

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Cambridge, UK

B.S. DeWitt and R. Stora, eds.
Les Houches, Session XL, 1983
Relativité, groupes et topologie II / Relativity, groups and topology II
© *Elsevier Science Publishers B.V., 1984*

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Contents

1. Introduction 336
2. Euclidean quantum field theory 338
3. Quantum fields in curved space 341
4. GUT era inflation 342
   4.1. The old Inflationary Model 345
   4.2. The new Inflationary Model 348
   4.3. Euclidean formulation of the Inflationary Model 350
   4.4. Growth of irregularities in the Inflationary Model 353
5. Quantum gravity 355
6. The conformal factor 361
7. The boundary conditions 362
8. Minisuperspace 364
   8.1. De Sitter space 364
   8.2. Our Universe? 369
9. Conclusion 376

References 379

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

1. Introduction

The subject of these lectures is quantum effects in cosmology. I shall deal first with situations in which the gravitational field can be treated as a classical, unquantized background on which the quantum matter fields propagate. This is the case with inflation at the GUT era. Nevertheless, the curvature of spacetime can have important effects on the behaviour of the quantum fields and on the development of long-range correlations. I shall then turn to the more exciting, and more speculative, question of the quantization of the gravitational field itself.

Quantum mechanics has been applied with great success to small systems (by small I mean anything from a two-particle system up to a white dwarf or neutron star). For such situations the Copenhagen Interpretation of quantum mechanics is adequate: the quantum system is supposed to interact with an external observer who is described by classical physics. Before the observation, the state of the system is a linear superposition of different eigenstates of the observable, but after the measurement of one of the eigenvalues of the observable, the system jumps to the corresponding eigenstate. This is the so-called "Collapse of the Wavefunction". However, the Copenhagen Interpretation is clearly no use if we want to consider the quantum mechanics of the whole universe, because, by its very definition, the universe includes everything of which we can have knowledge. There might be angels who could observe the Universe from outside but they could never communicate the results of their observations to us because, if they could, they would be part of the Universe and would be governed by the same laws, including quantum mechanics, as the rest of the universe. It is possible to overcome this difficulty and to give an interpretation of quantum mechanics which refers only to measurements *within* the system. This is usually called the "Many Worlds" Interpretation. I think it is a rather misleading name because it suggests that the universe is continually branching or dividing and people object that they do not feel themselves branching. In fact I think that the Many Worlds Interpretation simply involves the use of conditional probabilities, that is, the probability that A will occur given B. These conditional probabilities or correlations can be calculated by the standard rules of quantum mechanics. In my opinion there is no problem in applying quantum mechanics to the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

whole universe and my attitude to those who argue about the interpretation of quantum mechanics is reflected in a paraphrase of Goering's remark: "When I hear of Schroedinger's cat, I reach for my gun".

At the present time any quantum effects on cosmology are negligibly small: the large-scale structure of the universe is described by classical General Relativity to a very high degree of accuracy. However, the classical theory predicts that the universe must have had a singularity in the past at which the curvature of spacetime would have been infinite and at which it would not have been possible to apply the laws of physics (see Hawking and Ellis 1973). One would expect classical General Relativity to break down in very strong gravitational fields in which quantum gravitational effects become significant. The importance of the singularity theorems is therefore that they show that quantum effects were dominant at the beginning of the expansion of the universe. The initial conditions for the evolution of the universe during the classical epoch will be determined by quantum gravity. Thus, the specification of the boundary conditions of the universe is equivalent to giving a prescription for its quantum state. I shall put forward a proposal that the quantum state of the universe is defined by a path integral over all compact Euclidean (i.e. positive definite) metrics (Hawking 1982, Hartle and Hawking 1983). This incorporates the idea that the universe is completely self-contained: there is nothing outside. In other words, the boundary condition of the universe is that it has no boundary. I shall show that, at least in a simple model, this prescription for the quantum state gives a wavefunction that can be interpreted as a superposition of the wavefunctions corresponding to a number of different closed classical Friedmann models which each start off with a long period of exponential expansion and then go over to a matter-dominated phase. They expand to a very large radius and then recollapse and repeat the behaviour periodically. Such a wavefunction really does require the "Many Worlds" interpretation. There is no conflict with observation because the matrix elements of any observable between wavefunctions corresponding to different classical Friedmann models are very small.

The plan of these lectures is as follows:

Section 2 deals with the Euclidean approach to quantum field theory in flat space. In section 3, I extend these techniques to quantum fields on a curved background with the four-sphere, the Euclidean version of De Sitter space as a particular example. Section 4 deals with the GUT era

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

inflationary model. I show how this can be formulated in Euclidean terms and I discuss the growth of irregularities. In section 5, I describe how the gravitational field can be quantized by a Euclidean path integral and I define a wavefunction of three-geometries and matter field configurations which obeys the Wheeler–DeWitt equation. Section 6 deals with the path integral over the conformal factor. It is necessary to rotate the contour of integration in the complex plane in order to make the integral converge. In section 7, I discuss the boundary conditions on the metrics in the path integral. These determine the quantum state of the universe. I show that the only reasonable boundary conditions for cosmology are that one integrates over compact metrics. The consequences of this boundary condition are explored in section 8 where I consider simplified minisuperspace models in which most of the degrees of freedom are frozen out. A homogeneous isotropic model with a massive scalar field gives a wavefunction which could describe our universe. Section 9 is the Conclusion.

2. Euclidean quantum field theory

In this section, I shall motivate the Euclidean Approach to quantum gravity. Further details will be found in earlier reviews (Hawking 1979a, b, Eguchi et al. 1980). I shall start with ordinary quantum field theory in flat spacetime. This can be defined by a path integral over all field configurations in Minkowski space. For example, in the case of a scalar field ϕ, the amplitude to go from a field configuration $\phi(x_1)$ at time t_1 to be configuration $\phi(x_2)$ at time t_2 is

$$\langle \phi_2, t_2 | \phi_1, t_1 \rangle = \int \mathrm{d}[\phi] \exp \mathrm{i}(I[\phi]), \tag{2.1}$$

where $I[\phi]$ is the action of the field configuration $\phi(x, t)$, $\mathrm{d}[\phi]$ is a measure on the space of field configurations, and the path integral is taken over all field configurations which agree with the configuration $\phi(x_1)$ at time t_1 and with the configuration $\phi(x_2)$ at time t_2.

The quantum state of the scalar field can be described by a wavefunction $\Psi[\phi(x), t]$ which is a functional on the space of field configurations $\phi(x)$ at the time t. The wavefunction at time t_2 can be calculated from the wavefunction at time t_1:

$$\Psi[\phi_2, t_2] = \int \mathrm{d}[\phi_1] \langle \phi_2, t_2 | \phi_1, t_1 \rangle \Psi[\phi_1, t_1]. \tag{2.2}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

By considering times t_2 and t_1 that are very close to each other, one can derive the Schroedinger equation:

$$\frac{\partial \Psi[\phi(x), t]}{\partial t} = -iH\Psi[\phi(x), t], \tag{2.3}$$

where H is the Hamiltonian functional operator.

The integrand in eq. (2.1) is of unit modulus but rapidly varying phase. The path integral therefore oscillates but does not converge. In order to obtain a convergent path integral one therefore rotates the time coordinate 90° clockwise in the complex plane by defining a new "Euclidean" time coordinate $\tau = it$. The path integral now becomes.

$$\langle \phi_2, \tau_2 | \phi_1, \tau_1 \rangle = \int d[\phi] \exp(-\tilde{I}[\phi]), \tag{2.4}$$

where the path integral is now taken over all field configurations $\phi(x, \tau)$ which are real on the Euclidean space with coordinates (x, τ) and which approach the given values at τ_1 and τ_2. The quantity $\tilde{I}[\phi] = -iI[\phi]$ is the Euclidean action of the field configuration ϕ. In a well-behaved quantum field theory the Euclidean action $\tilde{I}$ is positive semi-definite and is zero only for the zero field configuration. The integrand in the Euclidean path integral (2.4) is therefore exponentially damped and the integral has a good chance of converging. One then analytically continues the amplitude (2.4) in the variable $\tau_2 - \tau_1$ anticlockwise in the complex t-plane back to real values.

The underlying reason why this Wick rotation in the complex t-plane makes the path integral converge is that in quantum mechanics a field with energy E has a time dependence of the form $\exp(-iEt)$. In a reasonable theory, $E \geqslant 0$. Thus a field configuration with a nonzero value of the energy will diverge at large positive imaginary t, i.e. at large real negative τ. The path integral over all fields which vanish in the past at positive imaginary values of t, i.e. at negative values of τ, and which vanish in the future at negative imaginary values of t, i.e. positive τ, gives the amplitude $\langle 0_+ | 0_- \rangle$ to go from the initial vacuum state $|0_-\rangle$ to the final vacuum state $\langle 0_+|$. Similarly the amplitude $\langle \phi, o | 0_- \rangle$ to go from the initial vacuum to the field configuration ϕ at time o, or in other words, the ground state wavefunction $\Psi_0[\phi, o]$, is given by a path integral over all fields ϕ in Euclidean space which match the given configuration at $\tau = o$ and which vanish at large negative τ. One can also introduce a term

$$\int J\phi \, d^3x \, d\tau \tag{2.5}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

into the Euclidean action, where $J(x,\tau)$ is a source term. Then the path integral over all field configurations ϕ in Euclidean space which vanish at large positive and negative values of the Euclidean time τ and at large spatial distances gives the amplitude to go from the initial vacuum to the final vacuum in the presence of the source J:

$$Z[J] = \langle 0_+|0_-\rangle_J = \int \mathrm{d}[\phi]\exp -\tilde{I}[\phi, J]. \tag{2.6}$$

One can functionally differentiate $Z[J]$ with respect to the source J at the points $x_1, \tau_1; x_2, \tau_2; \ldots$ and then set J to zero to obtain the Euclidean Green functions $G_E(x_1, \tau_1; x_2, \tau_2; \ldots)$. These will be symmetric in their arguments for bosonic fields and anti-symmetric for fermion fields. One can analytically continue G_E in the variables $\tau_1 - \tau_2$, $\tau_2 - \tau_3, \ldots$, to find the Green functions in Minkowski space. If one keeps the imaginary part of the time separations $t_1 - t_2$, $t_2 - t_3, \ldots$, negative, one obtains the Green function or vacuum expectation value with the operator ordering:

$$\langle 0|\phi(x_1, t_1), \phi(x_2, t_2), \ldots|0\rangle. \tag{2.7}$$

The expectation value of the field is given by

$$\langle \phi(x,\tau)\rangle = \frac{\delta W[J]}{J(x,\tau)}, \tag{2.8}$$

where $W[J] = -\log Z[J]$. One can invert eq. (2.8) to express the source J as a function of $\langle \phi(x,\tau)\rangle$. One can also define the effective action Γ as a function of $\langle \phi(x,\tau)\rangle$ by

$$\Gamma = W - \int J\langle \phi(x,\tau)\rangle \mathrm{d}^3x\mathrm{d}\tau. \tag{2.9}$$

The source J can then be expressed as

$$J = -\frac{\delta\Gamma}{\delta\langle\phi\rangle}. \tag{2.10}$$

Thus, in the absence of sources, the effective action is an extremum:

$$\frac{\delta\Gamma}{\delta\langle\phi\rangle} = 0. \tag{2.11}$$

The effective potential $V(\langle\phi\rangle)$ is the effective action per unit volume for constant expectation value $\langle\phi\rangle$. It will be of interest to us in connection with spontaneous symmetry breaking. If a Higgs field ϕ has an effective potential with a global minimum at zero, the expectation value $\langle\phi\rangle$ will

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

be zero and the gauge symmetry will be unbroken. However, if the global minimum occurs at a nonzero value ϕ_0, the lowest energy state will have $\langle\phi\rangle = \phi_0$ and the symmetry will be broken. Nevertheless, the system may be trapped in a local minimum of the effective potential at $\langle\phi\rangle = 0$. In this case the system will eventually tunnel quantum mechanically from the "false" vacuum at $\langle\phi\rangle = 0$ to the "true" vacuum at $\langle\phi\rangle = \phi_0$ (Coleman 1976). I shall return to symmetry breaking in connection with inflation at the GUT era in section 4.

3. Quantum fields in curved space

One can calculate path integrals for quantum fields in positive definite background metrics other than just flat Euclidean space. The simplest example is flat Euclidean space identified with a period β in the imaginary time coordinate τ. To see what the physical significance of this is, recall that the amplitude to go from a field configuration ϕ_1 at time t_1 to a configuration ϕ_2 at time t_2 can be expressed as a matrix element of the Hamiltonian between these two states:

$$\langle\phi_2, t_2|\phi_1, t_1\rangle = \langle\phi_2|\exp(-iH(t_2-t_1))|\phi_1\rangle. \tag{3.1}$$

If one rotates in the complex t-plane so that $t_2-t_1 = -i\beta$, puts the configuration ϕ_2 equal to ϕ_1 and sums over a complete set of configurations ϕ_n, the right-hand side of eq. (3.1) becomes the partition function $Z(\beta)$ at temperature $T = \beta^{-1}$:

$$Z(\beta) = \sum_n \exp(-\beta E_n), \tag{3.2}$$

where E_n is the energy of the configuration ϕ_n. By eq. (2.4), $Z(\beta)$ is also equal to the path integral over all fields ϕ on Euclidean space identified with period β in the imaginary time coordinate τ.

An example of a curved spacetime in which the substitution $\tau = it$ gives a positive definite metric that is periodic in imaginary time is the Schwarzschild solution

$$ds^2 = -(1-2Mr^{-1})dt^2 + (1-2Mr^{-1})^{-1}dr^2 + r^2 d\Omega^2. \tag{3.3}$$

Putting $\tau = it$ makes the metric positive definite for $r > 2M$. There is an apparent singularity in the metric at $r = 2M$ but, as in the Lorentzian case, it can be removed by defining a new coordinate,

$$x = 4M(1-2Mr^{-1})^{1/2}. \tag{3.4}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

The metric then takes the form

$$ds^2 = x^2\left(\frac{d\tau}{4M}\right)^2 + \left(\frac{r}{2M}\right)^4 dx^2 + r^2 d\Omega^2. \tag{3.5}$$

The apparent singularity in the x–τ plane at $x = 0$, $r = 2M$ is then just like the singularity in the metric in polar coordinates ρ, θ at $\rho = 0$,

$$ds^2 = \rho^2 d\theta^2 + d\rho^2. \tag{3.6}$$

Thus the metric (3.5) is regular at $x = 0$ if $\tau/4M$ is regarded as an angular coordinate and is identified with period 2π, i.e. if τ is identified with period $8\pi M$. This shows that quantum field theory defined on the Euclidean Schwarzschild metric corresponds naturally to a state at temperature $T = (8\pi M)^{-1}$.

Of more interest in cosmology is De Sitter space which is a solution of the Einstein equations with a cosmological constant Λ. This can be expressed in the form

$$ds^2 = -dt^2 + \left(\frac{1}{H^2}\cosh^2 Ht\right) d\Omega_3^2, \tag{3.7}$$

where $H^2 = \frac{1}{3}\Lambda$ and $d\Omega_3^2$ is the metric on a unit three-sphere. One can regard this as a homogeneous isotropic spacetime which collapses from infinite radius to a minimum radius of H^{-1} and then re-expands exponentially. Putting $\tau = it$ converts this to the positive definite metric on the four-sphere,

$$ds^2 = d\tau^2 + \left(\frac{1}{H^2}\cos^2 H\tau\right) d\Omega_3^2. \tag{3.8}$$

The orbits of $\partial/\partial\tau$ are great circles of longitude on the four-sphere. Thus, the metric (3.8) is periodic in the imaginary time coordinate τ with period $2\pi/H$. Thus, quantum field theory defined on the Euclidean De Sitter metric corresponds to a state at temperature $T = H/2\pi$.

4. GUT era inflation

The standard hot Big Bang model represents the universe by a Friedmann–Robertson–Walker metric,

$$ds^2 = -dt^2 + R(t)^2\left(\frac{dr^2}{1+kr^2} + r^2 d\Omega^2\right), \tag{4.1}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where $k = 1$ for a closed universe, $k = 0$ for a flat, open universe and $k = -1$ for a hyperbolic, open universe. The Universe is supposed to start off in thermal equilibrium at infinite temperature T at the initial singularity $t = 0$, $R(t) = 0$. As the universe expands, the temperature drops but most of the constituents of the model remain near to thermal equilibrium until the temperature falls to a few thousand degrees, at which point the electrons and light nuclei recombine and cease to be in equilibrium with the photons. This model has been very successful in accounting for the microwave background radiation and in predicting the abundances of the light elements like deuterium, helium and lithium. However, it has a number of features which are not explained but are merely assumed as initial conditions. Among these are:

(1) (The horizon problem)

The microwave radiation coming from different directions in the sky has the same temperature to one part in 10^3. One could understand this if the radiation came from regions which could have been in thermal equilibrium with each other. However, in the standard model, the radiation coming from directions which differ by more than a few degrees was last scattered in regions which had not been in causal contact with each other since the initial singularity (see fig. 1).

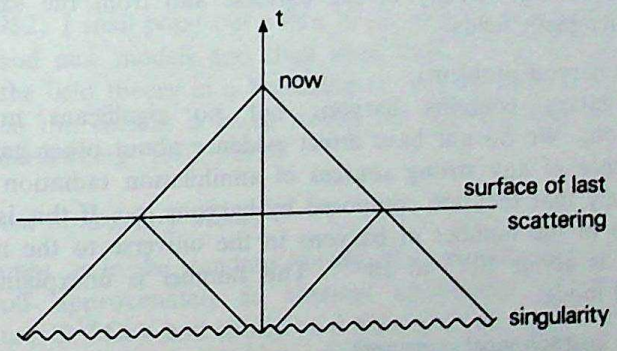


Fig. 1. A conformal diagram of the hot Big Bang model. The pasts of different regions on the surface of last scattering of the microwave background do not intersect.

(2) (The flatness problem)

The present average density ρ of the universe is within a factor of ten of the critical density ρ_c that characterizes the flat $k = 0$ model and which separates the models that recollapse from those which expand indefinitely. In order for $\Omega = \rho/\rho_c$ to be within a factor of ten of unity

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

now, it would have had to have been much closer at earlier times. For example, at $t = 1$ s, $|\Omega - 1| < 10^{-19}$.

(3) (The smoothness problem)

If the universe had been in thermal equilibrium at very early times, one would have expected statistical fluctuations of order $N^{1/2}$ in the number of particles in a given region, where N is the expected number. Because different regions were not in causal contact (the horizon problem), the fluctuations would have been frozen in and would have caused fluctuations in Ω much larger than is compatible with the isotropy of the microwave background.

(4) (The monopole problem)

Grand Unified Theories of the strong, weak and electromagnetic interactions predict the existence of superheavy magnetic monopoles with masses of the order of 10^{17} GeV. In the standard model, large numbers of these monopoles would be produced when the temperature fell below the Grand Unified Energy of about 10^{15} GeV because the directions of symmetry breaking would be different in different regions as they were not in causal contact with each other (cf. the horizon problem). The predicted number of monopoles would be about fourteen orders of magnitude greater than the upper limit that can be placed from the average density of the universe and from the existence of galactic magnetic fields.

(5) (The baryon problem)

Our galaxy contains baryons but no significant number of antibaryons. We do not have direct evidence about other galaxies but the absence of any strong sources of annihilation radiation makes it seem likely that they are composed by baryons too. If this is the case, the ratio of the number of baryons in the universe to the number of photons is about 10^{-8} to 10^{-10}. This number is unexplained in the standard model.

(6) (The cosmological constant)

In principle one could measure the cosmological constant by determining the Hubble Constant or rate of expansion $H = \dot{R}/R$ and the deceleration parameter $q_0 = -\ddot{R}/RH^2$. In practice, observational uncertainties prevent one from obtaining more than an upper bound $|q_0| < 5$ but this gives a very low upper limit,

$$\frac{|\Lambda|}{m_P^2} < 10^{-120},$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where m_P is the Planck mass. By contrast, there are a number of symmetry breaking effects which would be expected to give contributions to the effective cosmological constant of at least forty orders of magnitude greater than the observational upper limit. They would have to be incredibly finely balanced against each other and against the bare value of the cosmological constant.

I shall not discuss the cosmological constant problem here but it can be argued that, if one has some mechanism, such as a three-index antisymmetric tensor field or topological fluctuations, which can generate a cosmological constant of arbitrary magnitude, then zero is by far the most probable value for the total cosmological constant (Hawking 1983a, b). The baryon number of the universe can be explained by invoking Grand Unified Theories of the strong, weak and electromagnetic interactions. These allow processes in which baryon number is not conserved. If one also supposes that these processes are not invariant under CP, then the non-T invariant situation in the early universe will produce an excess of baryons over antibaryons of the order of 10^{-9}. Problems (1) to (4) remain however. In an attempt to provide an explanation of them, Guth (1981) proposed to so-called Inflationary Model of the Universe. I shall describe first the original version of the model and then the improved "new Inflationary Model" proposed by Linde (1982). I shall point out a flaw in the original formulation of both the old and new models and shall show how it can be corrected by defining the field theory in a Euclidean i.e. positive definite metric. The material in this section is based mainly on Hawking and Moss (1982, 1983) and Hawking (1982, 1983).

4.1. The old Inflationary Model

One assumes, as in the standard hot Big Bang model, that the universe started off approximately in thermal equilibrium at the Planck temperature 10^{19} GeV and cooled as it expanded. At temperatures much above the Grand Unified Energy of 10^{15} GeV, the effective potential $V(\Phi)$ of the Higgs scalar fields would have the form shown in fig. 2, with a minimum only at zero value of Φ.

The expectation value $\langle\Phi\rangle$ would therefore be zero and the Grand Unified Symmetry would be unbroken. As the universe expanded and the temperature fell, the form of $V(\Phi)$ would change and it would develop another local minimum for $\Phi \neq 0$. At a temperature T_c of order 10^{15} GeV, the other minimum would occur at the same value of the potential as the minimum at $\Phi = 0$ (see fig. 3). However, because of the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

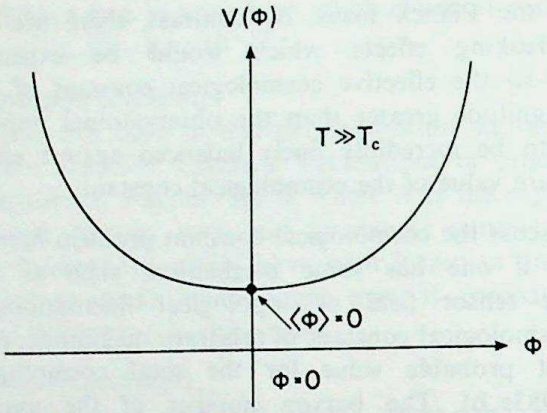


Fig. 2. The form of the effective potential of the Higgs scalar fields Φ at a temperature significantly above the Grand Unified Energy 10^{15}GeV. The expectation value $\langle\Phi\rangle$ lies at the minimum at $\Phi = 0$ and the Grand Unified Symmetry is unbroken.

potential barrier between the two minima, $\langle\Phi\rangle$ would remain at zero and the symmetry would remain unbroken. As the universe continued to expand, the temperature would fall to a low value and the other minimum would fall below that at $\Phi = 0$ (see fig. 4). The scalar field would however remain trapped in the metastable state $\langle\Phi\rangle = 0$. The energy density of the universe would come to be dominated by the nonzero value $V(0)$ of the potential in the metastable state. This would act like a cosmological constant and would cause the universe to expand

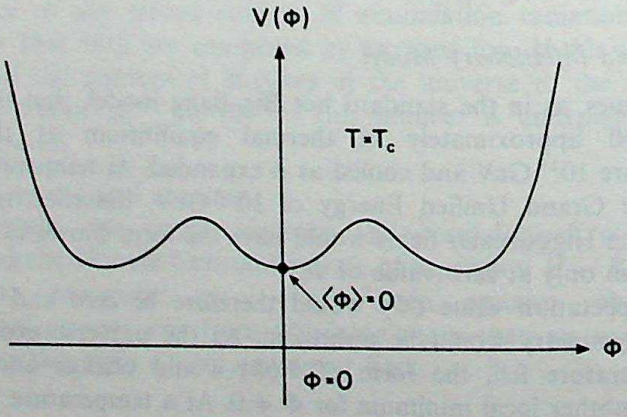


Fig. 3. At a critical temperature T_c of the order of 10^{15}GeV the effective potential has another local minimum for $\Phi \neq 0$. However, the expectation value $\langle\Phi\rangle$ remains at zero.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

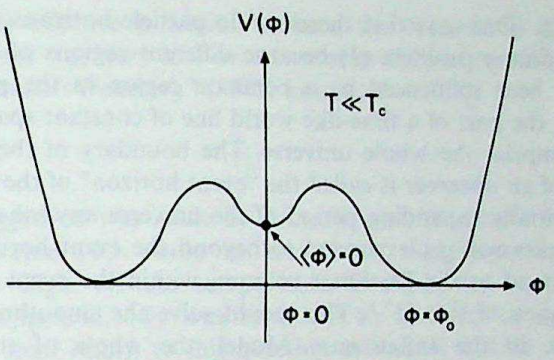


Fig. 4. At low temperatures the minimum at $\Phi \neq 0$ falls below that at $\Phi = 0$. However, the expectation value is trapped in a metastable state at $\Phi = 0$ by a potential barrier. Eventually, the expectation value tunnels quantum mechanically through the potential barrier to the broken symmetry state at $\Phi \neq 0$.

exponentially with $R(t) \propto \exp(Ht)$ where $3H^2 = 8\pi m_P^{-2} V(0)$. Eventually the scalar field would tunnel quantum mechanically through the barrier from $\langle \Phi \rangle = 0$ to the global minimum at $\langle \Phi \rangle = \Phi_0$ at which the Grand Unified Symmetry would be broken. The tunnelling was supposed to occur through the formation of a bubble of the broken symmetry phase surrounded by the unbroken symmetry phase (see fig. 5) (Coleman 1977). The bubble would appear as a result of a quantum fluctuation. It would expand and would collide with other bubbles to give a transition from the metastable state of unbroken symmetry to the stable state of broken symmetry. The value of the potential at $\langle \Phi \rangle = \Phi_0$ was supposed to be zero, so this would end the period of exponential expansion and the universe would continue to expand as in the standard model.

In an exponentially expanding universe, the pasts of any two points

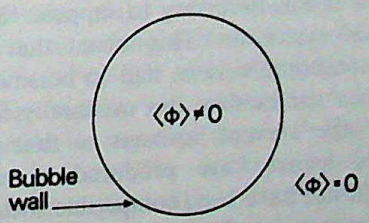


Fig. 5. The quantum tunnelling from the metastable state of unbroken symmetry at $\Phi = 0$ to the broken symmetry state at $\Phi = \Phi_0$ takes place by the formation of a bubble of the broken symmetry state surrounded by the unbroken symmetry state. The bubble walls accelerate outwards and eventually collide with the walls of other bubbles.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

will intersect. (One says that there are no particle horizons.) This would solve the horizon problem (1) because different regions of the universe could have been influenced by a common region in the past. On the other hand, the past of a time-like world line of constant spatial position will not comprise the whole universe. The boundary of the past of the world line of an observer is called the "event horizon" of the observer. In the exponentially expanding period of the universe any inhomogeneities or anisotropies will get expanded to beyond the event horizon and the metric will tend to the De Sitter solution within the event horizon, i.e. within a region of size H^{-1}. This would solve the smoothness problem (3) because, in the Inflationary Model, the whole of the presently observed universe would have evolved from the region within one event horizon in the exponential period. It would also solve the flatness problem (2) because any initial spatial curvature, that is, any departure from $\Omega = 1$, would be reduced to a very small value by the exponential expansion. Finally, the monopole problem (4) would also be solved because the number of monopoles would be of the order of the number of bubbles in the observed universe, and this number would be small. Indeed, as will be seen below, this is the main drawback of the original Inflationary Model.

4.2. The new Inflationary Model

During the exponential expanding phase, the temperature of the universe would fall to a very low value. When bubbles of the new broken symmetry phase formed, the vacuum energy $V(0)$ of the unbroken symmetry phase would go mostly into accelerating the walls of the bubbles. When the walls collided, this energy would be thermalized and would reheat the universe to a temperature near the Grand Unified Energy 10^{15} GeV. However, in order to solve the horizon, flatness and smoothness problems it was necessary to suppose that there was a long period of exponential expansion. This meant that the rate of bubble formation per unit spacetime volume had to be small compared to H^4. But in this case, when the bubbles did eventually form, they would be often beyond each other's event horizons so that only a few bubbles would collide. This would have produced a very inhomogeneous universe which would not have been compatible with the observations of the isotropy of the microwave background (Guth and Weinberg 1981, Hawking et al. 1982).

This difficulty in the original or "old" Inflationary Model was overcome by the suggestion of Linde (1982) that at low temperatures the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

barrier in the effective potential between the symmetric and broken symmetry states was very low or was absent altogether and that $V(\Phi)$ had the "Coleman–Weinberg" (1973) form at $\Phi = 0$, i.e.

$$\frac{\partial^2 V}{\partial \Phi^2} = 0. \tag{4.2}$$

A similar suggestion was made by Albrecht and Steinhardt (1982). The expectation value $\langle \Phi \rangle$ would be zero when the temperature of the universe was high. As the temperature fell and the barrier in the potential disappeared, $\langle \Phi \rangle$ was supposed to move away or fluctuate from zero (I shall return to this question in section 4.3). It would then start running down the slope of the effective potential (see fig. 6).

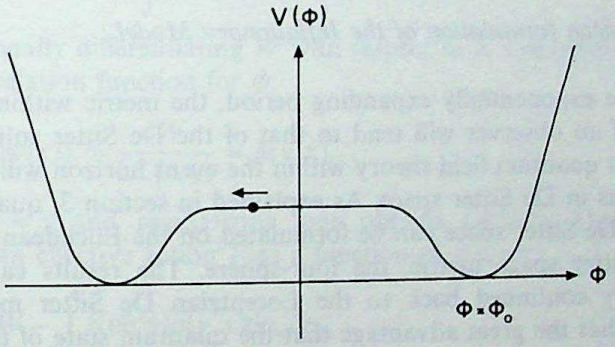


Fig. 6. In the new Inflationary Model there is supposed to be no potential barrier at low temperatures. The expectation value $\langle \Phi \rangle$ starts to run down-hill to the global minimum at Φ_0. Because the potential is assumed to be very flat near $\Phi = 0$, the initial rate of rundown is low and the universe expands exponentially by a large factor.

Because of the Coleman–Weinberg condition (4.2), it would move only slowly at first and the value of the potential would be almost unchanged. One could therefore have a long period of exponential expansion, as before. Eventually, when $\langle \Phi \rangle$ got near the global minimum at Φ_0 it would start to change rapidly and would oscillate about Φ_0. These oscillations of $\langle \Phi \rangle$ could be regarded as a very large number of Higgs particles in a coherent state. The universe would cease to expand exponentially and would expand like a matter-dominated Friedmann universe with $R(t) \propto t^{2/3}$. The Higgs particles would decay into light particles and would reheat the universe to a temperature near the Grand Unified Energy, 10^{15} GeV. The universe would then expand with $R(t) \propto t^{1/2}$. This "new" Inflationary Model has the advantages of the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

original model but manages to exit from the exponentially expanding period without creating large amounts of inhomogeneity. One would expect however, that the scalar field Φ would start running down the hill of the effective potential at different times in different places. This would give rise to a universe with a spectrum of adiabatic perturbations that was scale free in the sense that the relative amplitude $\delta\rho/\rho$ was the same on all length scales at the time that the length scale was equal to the expansion time H^{-1} (Hawking 1982, Guth and Pi 1982). Such a spectrum of perturbations could account for the formation of galaxies but the predicted amplitude, at least in the simplest Grand Unified Theories, was greater than unity rather than less than the upper limit of about 10^{-4} set by observations of the microwave background. I shall consider perturbations in more detail in section 4.4.

4.3. Euclidean formulation of the Inflationary Model

During the exponentially expanding period, the metric within the event horizon of an observer will tend to that of the De Sitter solution. This means that quantum field theory within the event horizon will be almost the same as in De Sitter space. As explained in section 3, quantum field theory in De Sitter space can be formulated on the Euclidean version of the De Sitter space metric, the four-sphere. The results can then be analytically continued back to the Lorentzian De Sitter metric. This approach has the great advantage that the quantum state of the fields is well defined by the condition that the path integral is taken over all field configurations that are regular on the four-sphere. This corresponds to a state at temperature $H/2\pi$. By contrast, if one works entirely in the Lorentzian metric, it is not clear what quantum state one should use or how one should define it.

There is a problem which arises when one tries to define quantum field theory on a compact manifold, like the four-sphere, rather than on a non-compact manifold of infinite volume, like flat Euclidean space. This is that, in a theory which is invariant under $\Phi \to -\Phi$, the expectation value $\langle\Phi\rangle$ is always zero in the limit that $J \to 0$, where J is the source coupled to Φ described in section 2. The reason is that the configurations Φ and $-\Phi$ are equally probable. This is true also in flat Euclidean space. However, in this case the amplitude to tunnel from Φ to $-\Phi$ is zero because the action of a field configuration which went from Φ at $\tau = -\infty$ to $-\Phi$ at $\tau = \infty$ would be infinite as the three-volume is infinite. This means that the expectation value $\langle\Phi\rangle$ can be nonzero even in the limit $J \to 0$. However, in a compact manifold, the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

amplitude to tunnel from Φ to $-\Phi$ will be nonzero so that $\langle\Phi\rangle$ will be zero when $J = 0$. One therefore cannot use the formalism described in section 2 in which the effective action Γ was defined aş a functional of the expectation value $\langle\Phi\rangle$.

One can overcome this problem by adding to the action a term

$$\int \Phi(x)K(x, y)\Phi(y)\,\mathrm{d}v_x\,\mathrm{d}v_y, \tag{4.4}$$

where $K(x, y)$ is a two-point function coupled to the Higgs field (Hawking and Moss 1983). One then defines $W(K)$ as before by integrating over all field configurations on the four-sphere:

$$\exp(-W(K)) = \int \mathrm{d}[\Phi]\exp(-\tilde{I}[\Phi]). \tag{4.5}$$

By functionally differentiating W with respect to K one obtains the two-point correlation function for Φ,

$$\langle\Phi(y)\Phi(x)\rangle = \frac{\delta W}{\delta K(x, y)} = \rho(y, x). \tag{4.6}$$

Taking $x = y$ and regularizing, one obtains $\langle\Phi^2(x)\rangle$. One can also calculate an effective action Γ as a function of $\rho(x, y)$:

$$\Gamma = W - \int K(x, y)\rho(y, x)\mathrm{d}v_x\,\mathrm{d}v_y. \tag{4.7}$$

One can evaluate this to one loop in the inflationary universe

$$\Gamma = \tfrac{1}{2}\mathrm{tr}[(-\Box + m^2)\rho - \log\rho] + \Gamma_0 + \int \mathrm{d}v_x \frac{3e^4}{32\pi^2}\rho_1^2 \log[M^{-2}(e^2\rho_1 + \tfrac{1}{2}H^2)], \tag{4.8}$$

where $\rho_1(x) = \rho(x, x)_{\mathrm{REN}} = \langle\Phi^2(x)\rangle$, e is the gauge coupling constant, m is the effective mass of the Higgs field including any coupling to the scalar curvature, M is a renormalization constant, and Γ_0 is the vacuum energy in the unbroken symmetry phase (Hawking and Moss 1983). The first term in (4.8) is the effective action of a free scalar field expressed as a function of ρ. The second term is a constant which is added to the action to make the effective cosmological constant zero in the broken symmetry phase in which $\langle\Phi^2\rangle = \Phi_0^2$. The last term is an effective Φ^4 interaction which is induced when one performs the path integral over the gauge field. Note the extra term in the logarithm which removes the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

usual infrared divergence at $\Phi = 0$. This arises because the four-sphere is compact. The derivative of the effective action with respect to ρ is the source term K,

$$\frac{\delta\Gamma}{\delta\rho(x,y)} = -K(y,x). \tag{4.9}$$

Thus, the equation of motion for ρ is

$$\frac{\delta\Gamma}{\delta\rho} = 0, \tag{4.10}$$

$$(-\Box + m^2)\rho(x,y) = \delta(x,y) - \frac{3e^4}{16\pi^2}\lambda\rho_1(x)\rho(x,y), \tag{4.11}$$

where λ is the logarithm in eq. (4.8) and the variation of the logarithm has been neglected.

In the hot early stages of the universe, the symmetry would have been unbroken and $\langle\Phi^2\rangle$ would have had the thermal value cT^2, where c is a numerical constant of order one. As the universe expanded, the temperature T would have dropped until the vacuum energy dominated over the thermal energy. The universe would then have entered an exponential expanding stage with

$$H^2 = \frac{m_\rho^2}{\pi\Gamma_0}. \tag{4.12}$$

As the temperature fell to the De Sitter value $H/2\pi$, $\langle\Phi^2\rangle$ would have fallen to the order of H^2 at some time $t = t_0$. The subsequent evolution of $\langle\Phi^2\rangle$ would have been governed by eq. (4.11). If $m^2 \ll H^2$, as is necessary for the new Inflationary Model, $\langle\Phi^2\rangle$ will initially increase linearly with time,

$$\langle\Phi^2\rangle \approx \frac{H^3}{4\pi^2}(t - t_0). \tag{4.13}$$

When $\langle\Phi^2\rangle$ becomes greater than $e^{-2}H^2$ at time $t = t_1 = t_0 + (4\pi e^2 H)^{-1}$, the nonlinear terms in eq. (4.11) will become important. The solution is then

$$\langle\Phi^2\rangle = \rho_1(t) \approx \frac{8\pi^2 H}{e^4\lambda(t_2 - t)}, \tag{4.14}$$

where $t_2 - t_1 \approx \lambda^{-0.5}e^{-2}H^{-1}$. Equation (4.14) holds until time t_3, where $t_3 - t_2 = H^{-1}$, i.e. until about one expansion time before t_2. After this

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$\langle\Phi^2\rangle$ will increase to, and oscillate about, its symmetry breaking value of $e^{-2}M^2$ in a time short compared to H^{-1}. When $\langle\Phi^2\rangle$ has this large value, the tunnelling between Φ and $-\Phi$ can be neglected and one can treat the system as having a large expectation value $\langle\Phi\rangle$, as before. The oscillations of $\langle\Phi^2\rangle$ can again be regarded as a large number of scalar particles in a coherent state. They will decay and reheat the universe to a temperature near the Grand Unified Temperature. The subsequent evolution of the universe will be the same as in the standard hot Big Bang model.

4.4. The growth of irregularities in the Inflationary Model

At the beginning of the inflationary period at $t = t_0$, the two-point correlation function $\rho(x, y) = \langle\Phi(x)\Phi(y)\rangle$ will be of order H^2. One can calculate it for points x and y on later surfaces of constant time. For $t < t_3$,

$$\rho(x, y) = \rho_1(t) + \rho_0(x, y), \tag{4.15}$$

where $\rho_0(x, y)$ is the analytic continuation of the Green function on the four-sphere with the $l = 0$ mode projected out, i.e. $\rho_0(x, y)$ is the solution of

$$(-\Box + m^2)\rho_0(x, y) = \delta(x, y) - \frac{3H^4}{8\pi^2}, \tag{4.16}$$

$$\rho_0(x, y) \approx -\frac{H^2}{4\pi^2}\log(H|x-y|), \quad |x-y| \gg H^{-1}, \tag{4.17}$$

where $|x-y|$ is the proper distance between x and y on the surface of constant t.

The expectation value $\langle\Phi^2\rangle$ will be constant on surfaces of constant time but there will be spatial fluctuations in Φ^2 which will be described by the two-point function

$$\langle\Phi^2(x)\Phi^2(y)\rangle = (\rho_1(t))^2 + 2\rho_1(t)\rho_0(x, y). \tag{4.18}$$

It is convenient to change to a new time coordinate, $\tilde{t} = t + \delta t$, which is constant on surfaces of constant Φ^2:

$$\delta t(x) - \delta t(y) = \frac{\Phi^2(y) - \Phi^2(x)}{\dot{\rho}_1(t)}. \tag{4.19}$$

The energy–momentum tensor will be nearly constant on these surfaces because it is dominated by the potential term in the effective action and

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

this is a function of Φ^2. However, there will be fluctuations δH in the rate of expansion H:

$$\delta H = \tfrac{1}{3}\Delta\delta t. \tag{4.20}$$

These fluctuations will be characterized by the two-point function

$$\langle \delta H(x)\delta H(y)\rangle = \frac{\rho_1 \nabla_x^2 \nabla_y^2 \rho_0(x,y)}{9(\dot{\rho}_1)^2}. \tag{4.21}$$

Before time t_3 one can neglect the gravitational effects of the fluctuations in the Higgs field and treat the spacetime as De Sitter space. After t_3 the universe will reheat and become radiation dominated in a time short compared to the expansion time H^{-1}. The details of the reheating do not matter apart from the fact that is occurs simultaneously in the $\bar{t}$ coordinate. Local conservation of energy will then imply that the radiation-dominated universe starts off with a uniform energy density but with fluctuations in the rate of expansion given by (4.21) with $t = t_3$:

$$\langle \delta H(x)\delta H(y)\rangle = \frac{\lambda}{9\pi^2}\left(\frac{e^2}{4\pi}\right)^2 H^{-2}|x-y|^{-4}. \tag{4.22}$$

Comoving regions that are separated by more than the horizon distance will evolve like independent Friedmann universes with the same initial density but different initial rate of expansion. Equation (4.22) will continue to hold until $|x-y| = H^{-1}$, i.e. until x and y are separated by one expansion distance. The amplitude $\delta H/H$ of fluctuations at this time will be

$$\left(\frac{\lambda}{9\pi^2}\right)^{1/2}\frac{e^2}{4\pi}, \tag{4.23}$$

independent of the scale of the fluctuation. One therefore has an exactly scale-free spectrum of perturbations as suggested by Harrison (1970) and Zel'dovich (1972). The amplitude given by (4.23) is smaller than previous estimates by Hawking (1982), Guth and Pi (1982) and Bardeen et al. (1982). For standard GUTs for which $e^2/4\pi \approx 1/40$, it would still be inconsistent with the upper limit of about 10^{-4} given by observations of the isotropy of the microwave background. However, in the case of supersymmetric GUTs, the effective self-coupling of the Higgs field would be smaller because the fermion loops contribute with the opposite sign to the vector loops. The predicted fluctuations might therefore be in agreement with the universe that we see around us.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

5. Quantum gravity

In the previous sections I have considered quantum field theory on a fixed unquantized curved space background. This is a good approximation at the GUT era because the typical energies are 10^{-4} of the Planck energy and so quantum fluctuations of the gravitational field are unimportant. However, if we want to understand the origins of the universe, one has to go back to the Planck era and include the gravitational field in the path integral. In the cases of the black hole and De Sitter metrics considered in section 3, it was possible to rotate the time coordinate to imaginary values and to pass from a Lorentzian metric (signature $-+++$) to a Euclidean metric (signature $++++$). However, in the case of a general Lorentzian metric, it is not possible to find a section in the complexified manifold on which the metric is real and Euclidean. Similarly, a general Euclidean metric will not have a section on which the metric is real and Lorentzian. One therefore has to decide whether to take the path integral to be over all Lorentzian metrics or over all Euclidean metrics. I shall take the point of view that these two are equivalent in the sense of contour integration. I shall show that the Wheeler–DeWitt equation, the analogue of the Schroedinger equation, is the same in the two cases but that the form of the wavefunction determines whether the quantum state corresponds to a Lorentzian or a Euclidean geometry in the classical limit. I shall use a path integral over all Euclidean (i.e. positive definite) metrics because this allows a simple formulation of the boundary conditions that I wish to impose. It also makes the path integral non-oscillatory although there is still a problem with convergence because the Euclidean action of the gravitational field is not positive definite (Gibbons et al. 1978). I shall show how to deal with this in section 6.

This section and sections 7 and 8.1 are based on Hartle and Hawking (1983). I shall assume that the probability for a four-metric $g_{\mu\nu}$ and a matter field configuration ϕ is proportional to

$$\exp -\tilde{I}[g_{\mu\nu}, \phi], \tag{5.1}$$

where $\tilde{I}$ is the Euclidean action

$$\tilde{I}[g_{\mu\nu}, \phi] = \frac{m_{\mathrm{P}}^2}{16\pi}\left[-\int_{\partial M} 2Kh^{1/2}\,\mathrm{d}^3x - \int_M \left(R - 2\Lambda + \frac{16\pi}{m_{\mathrm{P}}^2} L(g_{\mu\nu}, \phi)\right) g^{1/2}\,\mathrm{d}^4x\right], \tag{5.2}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where h_{ij} is the three-metric on the boundary ∂M and K is the trace of the second fundamental form of the boundary. The surface term in the action is necessary because the curvature scalar R contains second derivatives of the metric (York 1972, Gibbons and Hawking 1977, Hawking 1979a). The physics of the universe is governed by probabilities of the form (5.1) for all four-metrics $g_{\mu\nu}$ and matter field configurations belonging to a certain class C. The specification of the class C determines the quantum state of the universe. This will be considered further in section 7.

In practice, one is normally interested in the probability, not of the entire four-metric, but of a more restricted set of observables. Such a probability can be derived from the basic probability (5.1) by integrating over the unobserved quantities. For example, one would like to define a wavefunction which describes the quantum state of the universe "at one time" in a manner similar to that in which one defined a wavefunction for a scalar field as a functional of the field configurations on a surface of constant time. The analogue of a surface of constant time is a three-dimensional submanifold S. One could therefore define the wavefunction as a functional of the metric and matter field configurations of S. In fact, only the components h_{ij} of the metric that lie in the three-manifold S have any invariant meaning because the other components can be given any value by a coordinate transformation (gauge transformation) which leaves S unchanged. Similarly, there is no invariant way to specify the time coordinate of the surface S. In fact, the three-metric h_{ij} determines where S fits into the four-manifold M with metric $g_{\mu\nu}$. The wave function Ψ is therefore a functional only of the three-metric h_{ij} and the matter field configurations ϕ on S. In the case of noncompact asymptotically flat three-manifolds S, the three-metric h_{ij} does not uniquely specify the position of the surface at infinity. In this case, the wavefunction depends also on the time coordinate t of S at infinity. However, it does not seem that the universe is asymptotically flat, though I shall return to this point in section 7. I shall therefore assume that the wavefunction $\Psi[h_{ij}, \phi]$ is a functional only of the three-metric and the matter field configurations. It is given by a path integral over all Euclidean four-metrics $g_{\mu\nu}$ and matter field configurations belonging to the class C which have S as their boundary ∂M and which agree with the given three-metric h_{ij} and matter field configuration on S (see fig. 7):

$$\Psi[h_{ij}, \phi] = \int_C \mathrm{d}[g_{\mu\nu}]\mathrm{d}[\phi]\exp -\tilde{I}[g_{\mu\nu}, \phi]. \tag{5.3}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

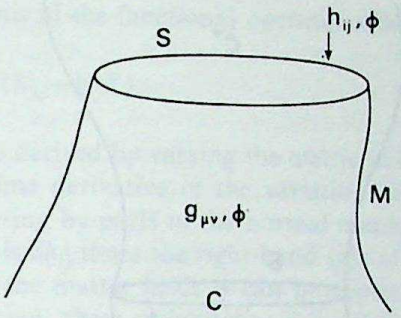


Fig. 7. The amplitude or wavefunction $\Psi[h_{ij}, \phi]$ is given by integrating over all four-geometries and matter field configurations belonging to the class C, which are bounded by the three-surface S and which agree with the given three-metric h_{ij} and matter field configuration ϕ on S.

Note that one integrates only over the four-geometries on one side of S. This gives one an *amplitude* Ψ for the given three-geometry h_{ij} and the matter field configuration ϕ. The *probability* $\Psi\Psi^*$ of finding a three-manifold with the given metric h_{ij} and matter field configuration ϕ can be represented by a path integral over four-metrics $g_{\mu\nu}$ on both sides of S (see fig. 8).

The wavefunction $\Psi[h_{ij}, \phi]$ obeys a functional differential equation, the Wheeler–DeWitt equation, which is the analogue of the Schroedinger equation. This can be derived directly from the path integral definition (5.3) as follows. In the neighbourhood of the surface S, one can introduce a time coordinate τ which is constant on S so that the metric takes the standard $3+1$ form:

$$ds^2 = (N^2 - N_i N^i)d\tau^2 + 2N_i\,dx^i\,d\tau + h_{ij}\,dx^i\,dx^j. \tag{5.4}$$

The action then becomes

$$\tilde{I}[g_{\mu\nu}, \phi] = \frac{m_P^2}{16\pi}\int d^3x\,d\tau\,h^{1/2}N[K_{ij}K^{ij} - K^2 - {}^3R(h^{ij}) + 2\Lambda - 16\pi m_P^{-2}L(\phi)], \tag{5.5}$$

where K_{ij} is the second fundamental form

$$K_{ij} = \frac{1}{N}\left[\frac{1}{2}\frac{\partial h_{ij}}{\partial \tau} + N_{(i|j)}\right), \tag{5.6}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

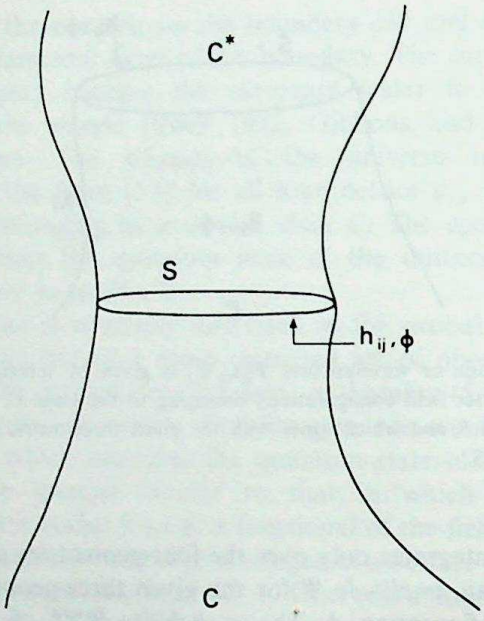


Fig. 8. The probability $\Psi\Psi^*$ of finding a three-surface S with the given three-metric and matter field configuration is given by integrating over all four-geometries and matter field configurations on both sides of S.

and a stroke and ${}^3R$ denote the covariant derivative and scalar curvature constructed from the three-metric h_{ij}.

The functional integral defining the wavefunction contains an integral over N. By varying N at the surface one pushes it forward or backward in time. Since the wavefunction does not depend on time one must have

$$0 = \int_C \mathrm{d}[g_{\mu\nu}]\mathrm{d}[\phi]\left(\frac{\delta \tilde{I}}{\delta N}\right)\exp(-\tilde{I}[g_{\mu\nu},\phi]). \tag{5.7}$$

Classicially the field equation $H \equiv \delta\tilde{I}/\delta N = 0$ is the Hamiltonian constraint for general relativity. It is:

$$H = -h^{1/2}(K^{ij}K_{ij} - K^2 + {}^3R - 2\Lambda + 8\pi m_{\mathrm{P}}^{-2}T_{\mathrm{nn}}) = 0, \tag{5.8}$$

where T_{nn} is the Euclidean stress energy tensor of the matter field projected in the direction normal to the surface. One can substitute the expression (5.8) for $\delta\tilde{I}/\delta N$ in (5.7). The second fundamental form K_{ij} can

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

be expressed in terms of the functional operator $\delta/\delta h_{ij}$:

$$\frac{\delta \tilde{I}}{\delta h_{ij}} = \frac{m_P}{16\pi} h^{1/2}(K_{ij} - h_{ij}K). \tag{5.9}$$

This identity can be derived by varying the metric in the definition of the action (5.2). The time derivative of the variation of the metric can be removed by integrating by parts in the normal manner. This produces a surface term which is δh_{ij} times the right-hand side of (5.9). Similarly, the time derivatives of the matter fields ϕ can be expressed in terms of the operator $\delta/\delta\phi$. Using these substitutions in (5.7), one obtains the Wheeler–DeWitt equation,

$$\left[-G_{ijkl} \frac{\delta^2}{\delta h_{ij} \delta h_{kl}} - h^{1/2}\left({}^3R(h) - 2\Lambda + 8\pi m_P^{-2} T_{nn}\left(\frac{\delta}{\delta\phi}, \phi\right)\right)\right] \Psi[h_{ij}, \phi] = 0, \tag{5.10}$$

where G_{ijkl} is the metric on superspace, the space of all three-metrics h_{ij}:

$$G_{ijkl} = \tfrac{1}{2} h^{1/2}(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}). \tag{5.11}$$

There is a question of the factor ordering in the functional differentiation but it will not be important in the situations that I shall consider. The Wheeler–DeWitt equation is the same as the equation one would have obtained had one started out with a path integral over Lorentzian metrics. One can also consider variations within the three-surface S generated by the "Shift" vector N_i:

$$0 = \int_C \frac{\delta \tilde{I}}{\delta N_i} \exp(-\tilde{I}[g_{\mu\nu}, \phi]). \tag{5.12}$$

This equation implies that the wavefunction is invariant under diffeomorphisms, i.e. that Ψ is a functional of the three-geometry and not of the particular three-metric h_{ij}.

The metric G_{ijkl} on superspace, the space of three-geometries, has a signature $(-+++++)$. The Wheeler–DeWitt equation can therefore be thought of as a hyperbolic equation on superspace with $h^{1/2}$ as the time coordinate. I shall show that the semiclassical approximation to the path integral gives the boundary conditions for the Wheeler–DeWitt equation at small $h^{1/2}$. The Wheeler–DeWitt equation can then be solved to give the values of the wavefunction at larger values of $h^{1/2}$. In a Lorentzian metric, $h^{1/2} = 0$ would be a singularity. However, this is

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

not necessarily the case in a Euclidean metric as can be seen by considering the example of a four-sphere of radius R embedded in flat five-dimensional space: a surface $|x^5| < R$ intersects the four-sphere in a three-sphere of nonzero radius. However, when $|x^5| = R$, the three-sphere shrinks to zero radius but there is no singularity of the four-geometry. Indeed, $h^{1/2}$ has to go to zero if the topology of the three-surfaces is to change.

By this construction, the wavefunction $\Psi[h_{ij}, \phi]$ vanishes for three-metrics h_{ij} which are not positive definite, i.e. for which the determinant h of the metric is negative. This means that h_{ij} is not a quantum observable with an unrestricted range and that in many cases it is more convenient to replace it by $\tilde{h}_{ij}$, the three-metric up to a conformal factor, and $Km_P^2/12\pi$, the momentum conjugate to $h^{1/2}$. One can then define a wavefunction Φ in this representation by

$$\Phi[\tilde{h}_{ij}, K, \phi] = \int_C \mathrm{d}[g_{\mu\nu}] \exp - I^{\mathrm{k}}[g_{\mu\nu}, \phi], \tag{5.13}$$

where $I^{\mathrm{k}}[g_{\mu\nu}, \phi]$ is the action appropriate to the situation in which K and $\tilde{h}_{ij}$ are kept fixed on the boundary rather than h_{ij}. It differs from the action $\bar{I}[g_{\mu\nu}, \phi]$ in that the surface term has a different coefficient, $\frac{2}{3}K$ rather than $2K$. The wavefunction $\Phi[h_{ij}, K, \phi]$ is the Laplace transform of $\Psi[h_{ij}, \phi]$,

$$\Phi[\tilde{h}_{ij}, K, \phi] = \int_0^\infty \mathrm{d}[h^{1/2}] \exp\left(- \frac{m_P^2}{12\pi} \int \mathrm{d}^3x h^{1/2} K \right) \Psi[h_{ij}, \phi]. \tag{5.14}$$

Similarly, $\Psi[h_{ij}, \phi]$ is the inverse Laplace transform of $\Phi[\tilde{h}_{ij}, K, \phi]$,

$$\Psi[h_{ij}, \phi] = \int_\Gamma \mathrm{d}\left[\frac{m_P^2}{24\pi \mathrm{i}} K \right] \exp\left(\frac{m_P^2}{12\pi} \int \mathrm{d}^3x h^{1/2} K \right) \Phi[\tilde{h}_{ij}, K, \phi], \tag{5.15}$$

where the contour Γ runs from $-\mathrm{i}\infty$ to $+\mathrm{i}\infty$ to the right of any singularities of $\Phi[\tilde{h}_{ij}, K, \phi]$ in the complex K plane. Providing that $\Phi[\tilde{h}_{ij}, K, \phi]$ does not diverge exponentially for large $\mathrm{Re}\, K$, this choice of contour will ensure that $\Psi[h_{ij}, \phi] = 0$ for $h^{1/2} < 0$ because one can close the contour in the right half K plane.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

6. The conformal factor

The Euclidean gravitation action is

$$\tilde{I}[g] = -\frac{m_P^2}{16\pi}\left\{\int_m (R-2\Lambda)g^{1/2}\,d^4x + \int_{\partial m} 2Kh^{1/2}\,d^3x\right\}. \tag{6.1}$$

Under a conformal transformation $\tilde{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$,

$$\tilde{I}[\tilde{g}] = -\frac{m_P^2}{16\pi}\left\{\int_m (R\Omega^2 + 6(\nabla\Omega)^2 - 2\Lambda\Omega^4)g^{1/2}\,d^4x + \int 2\Omega^2 Kh^{1/2}\,d^3x\right\}. \tag{6.2}$$

One can make the action arbitrarily negative by choosing a conformal factor which is rapidly varying. This behaviour does not indicate a real instability. In fact the fluctuations of the conformal factor are cancelled out by the ghosts. This lack of positivity is closely related to the fact that gravitational potential energy is negative because gravity is attractive. Nevertheless, in the classical theory the total energy, rest mass energy plus potential energy, is always positive for an isolated system. What happens is when the negative potential energy becomes large and threatens to overwhelm the rest mass energy, a horizon forms and the system collapses to a black hole. The energy of a black hole is always positive (Gibbons et al. 1983).

What happens in the quantum theory is rather similar. One can divide the space of all metrics $g_{\mu\nu}$ into equivalence classes under conformal transformations. In each equivalence class one can rotate the contours of integration over the conformal factor at each point so that it is parallel to the imaginary axis. This makes the kinetic term

$$-\frac{m_P^2}{16\pi}\int_m 6(\nabla\Omega)^2 g^{1/2}\,d^4x \tag{6.3}$$

positive definite. The cosmological term

$$\frac{m_P^2}{16\pi}\int_m 2\Lambda\Omega^4 g^{1/2}\,d^4x \tag{6.4}$$

will be positive provided that $\Lambda > 0$ and will dominate the term

$$-\frac{m_P^2}{16\pi}\int_m R\Omega^2 g^{1/2}\,d^4x. \tag{6.5}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Thus the path integral over the conformal factor for each conformal equivalence class of metrics will converge with this prescription. The positive action theorem indicates that the path integral over different conformal equivalence classes of metrics will also converge.

For certain conformal equivalence classes of metrics there will be a real function $\tilde{\Omega}(x)$ such that $\tilde{I}[\tilde{\Omega}^2 g]$ is a local maximum. In this case one can distort the contour of the Ω integration at each point x so that it crosses the real Ω axis at $\tilde{\Omega}(x)$. In this case $\tilde{\Omega}(x)$ will be the dominant saddle point in the Ω integration and the path integral over the conformal factor will be of the order of $e^{-\tilde{I}[\tilde{\Omega}^2 g]}$.

Note that there may be more than one extremum of $\tilde{I}[\Omega^2 g]$ on the real Ω axis but the dominant contribution comes from the one with the greatest action, and not the least action as one might have expected. Indeed, as I have shown above, $\tilde{I}[\Omega^2 g]$ does not have any minimum on the real axis. In section 8.1, I shall show an example in which there is more than one extremum of $\tilde{I}[\Omega^2 g]$ on the real Ω axis and I will give an independent argument to show that the dominant contribution to the path integral comes from the extremum which has the greatest value of the action.

For other equivalence classes of metrics there will not be any real $\tilde{\Omega}(x)$ which makes $\tilde{I}[\tilde{\Omega}^2 g]$ a local maximum. In these cases the dominant contribution to the path integral over Ω will come from complex saddle points $\tilde{\Omega}(x)$. At these saddle points $\tilde{I}[\tilde{\Omega}^2 g]$ will be complex. The saddle-point contribution of order $e^{-\tilde{I}[\tilde{\Omega}^2 g]}$ will have an oscillating phase and will tend to cancel out with the contributions from neighbouring conformal equivalence classes of metrics unless they satisfy the classical field equations. Thus, one would expect these conformal equivalence classes of metrics to produce a rapidly oscillating path integral and to be damped unless they contain a complex solution of the field equations. I shall show an example of this in section 8.1.

7. The boundary conditions

In this section I shall discuss the boundary conditions for the class C of metrics that determines the quantum state of the universe. In the case of positive cosmological constant Λ any nonsingular, geodesically complete positive definite metric which is a solution of the field equations is necessarily compact. Its four-volume is bounded by the solution with highest degree of symmetry, which is a four-sphere of radius $\sqrt{3/\Lambda}$. Thus, in the case of positive Λ it seems clear that one should do the path

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

integral over compact metrics. That is to say, the wavefunction $\Psi[h_{ij}\phi]$ should be defined by a path integral over all compact metrics and scalar field configurations which have the surface S as the boundary on which they match the given three-metric h_{ij} and matter field configuration ϕ.

In the case of zero or negative Λ there are solutions of the field equations which are not compact and which have infinite four-volume. The solutions with the greatest symmetry are flat Euclidean space and Euclidean anti-De Sitter (ADS) respectively. (Euclidean ADS is the homogeneous space O(4, 1)/O(4)). Thus one might think that one should choose the class C to be asymptotically Euclidean metrics or asymptotically ADS respectively. This is appropriate if one is considering scattering problems in which one sends waves in from infinity and then measures what comes back to infinity. In this situation one is interested only in metrics which are connected to infinity. If there were any disconnected compact parts they would not contribute to the scattering of particles coming from infinity. However, in cosmology, one is interested in measurements in one's immediate neighbourhood, not at infinity. If one adopted asymptotically Euclidean or ADS boundary conditions, one would also have contributions from disconnected metrics which consist of a compact metric and an asymptotically Euclidean or ADS metric. One cannot exclude disconnected metrics from the path integral because they can be approximated by connected metrics in which the different components are joined by thin tubes with negligible action. Similarly, topologically nontrivial metrics can be approximated arbitrarily closely by topologically trivial metrics. It turns out that the disconnected metrics tend to dominate the path integral because the compact metrics can have negative action. Thus, even if one adopted asymptotically Euclidean or asymptotically ADS boundary conditions, the wavefunction that one would derive would be almost the same as that which one would obtain from compact metrics. I shall therefore assume that the class C should consist of compact metrics. It should be emphasized that this is a *proposal* for the boundary conditions of the universe. One cannot derive it from other principles. One can merely show that it is reasonable, and what could be more reasonable than the boundary condition that the universe has no boundary? The ultimate test, however, is to show that it predicts a universe like the one that we live in. I shall attempt to do this for a simple model in section 8.2.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

8. Minisuperspace

In order to investigate whether the boundary conditions proposed in section 7 correspond to the quantum state of the universe that we live in, one would like to calculate the wavefunction $\Psi[h_{ij}, \phi]$ and see if it can be interpreted as predicting what we observe. This is equivalent to finding a solution of the Wheeler–DeWitt equation obeying certain boundary conditions. The Wheeler–DeWitt equation is a second-order hyperbolic differential equation on an infinite-dimensional manifold, superspace. We do not know how to solve such an equation but we can hope to get some idea of the nature of the solution by considering the equation restricted to a finite-dimensional submanifold called minisuperspace. In other words, one restricts the infinite number of degrees of freedom of the gravitational and matter fields down to a finite number. The Wheeler–DeWitt equation then becomes a wave equation on a finite-dimensional manifold and can be solved by standard techniques. The boundary conditions for the Wheeler–DeWitt equation can be obtained from the semiclassical approximation to the path integral, i.e. by expressing the wavefunction in the form

$$\Psi[h_{ij}, \phi] = N_0 \sum_i A_i \exp - B_i, \tag{8.1}$$

where N_0 is a normalization constant, and B_i are the actions of classical solutions of the field equations which are compact and which have the given three-metric h_{ij} and matter field configuration ϕ on the boundary. The prefactors A_i are given by the determinants of small fluctuations about the classical solutions. I shall not bother with them but only with the more important quantities B_i which appear in the exponential.

8.1. De Sitter space

The first example of a minisuperspace that I shall consider is that for a homogeneous isotropic universe with no matter fields but with a positive cosmological constant Λ. Even though this is a very simple example, it will be useful because it will indicate how one should interpret the wavefunction. I shall show that it corresponds in the classical limit with the classical solution of highest symmetry, De Sitter space.

The metric of a homogeneous isotropic Euclidean space can be written in the form

$$ds^2 = \sigma^2[N^2(\tau)d\tau^2 + a^2(\tau)d\Omega_3^2], \tag{8.2}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where $N(\tau)$ is the lapse function, $\sigma^2 = 2/3\pi m_p^2$ and $d\Omega_3^2$ is the metric on a three-sphere of unit radius. The action is

$$\tilde{I} = \tfrac{1}{2}\int d\tau \left(\frac{N}{a}\right)\left[-\left\{\frac{a}{N}\frac{da}{d\tau}\right\}^2 - a^2 + \lambda a^4\right], \tag{8.3}$$

where $\lambda = \sigma^2 \Lambda/3$. The Wheeler–DeWitt equation is

$$\frac{1}{2}\left[\frac{1}{a^p}\frac{\partial}{\partial a}\left(a^p \frac{\partial}{\partial a}\right) - a^2 + \lambda a^4\right]\Psi[a] = 0. \tag{8.4}$$

The index p represents some, but not all, of the ambiguity in the factor ordering in the Wheeler–DeWitt equation. Its value will not make a great difference to the nature of the solution.

The wavefunction $\Psi(a)$ is given by a path integral over all compact metrics of the form (8.2) which are bounded by a three-sphere of radius a. These metrics are all conformally flat. The semiclassical approximation to the path integral will be given by the solution of the classical field equations with a Λ term. If $a < H^{-1}$ where $H^2 = \lambda$, there are two solutions. They correspond to the two regions into which a three-sphere of radius a divides a four-sphere of radius H^{-1} (see fig. 9). The actions of these metrics are

$$\tilde{I}_{\pm}(a) = -\frac{1}{3H^2}[1 \pm (1 - H^2a^2)^{3/2}], \tag{8.5}$$

where the plus sign is the action that corresponds to filling in the three-sphere with more than half the four-sphere and the minus sign is the

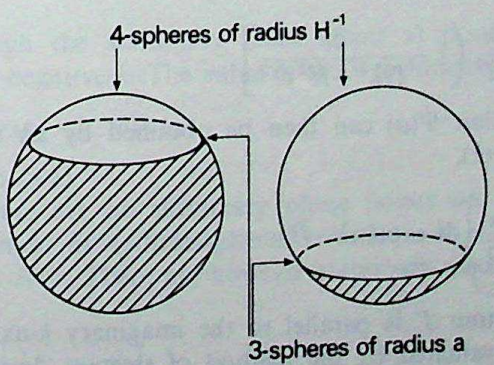


Fig. 9. There are two solutions of the field equations for a compact four-geometry which is bounded by a three-sphere of radius $a < H^{-1}$. They correspond to filling in the three-sphere with more or with less than half a four-sphere of radius H^{-1}.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

action for less than half the four-sphere. According to the discussion in section 6, the dominant contribution to the path integral will come from the greater of these conformally related actions, i.e. from $\tilde{I}_-(a)$. I shall derive the same result in another way below. The semiclassical approximation to the wavefunction is therefore

$$\Psi(a) \approx N_0 \exp\left(\frac{1}{3H^2}[1-(1-H^2a^2)^{3/2}]\right), \quad Ha < 1$$

$$\approx N_0 \exp(\tfrac{1}{2}a^2), \quad Ha \ll 1. \tag{8.6}$$

If $a > H^{-1}$, there is no real solution of the field equations because a three-sphere of radius a cannot be fitted into a four-sphere of radius H^{-1}. The dominant contribution to the path integral over the conformal factor will therefore come from a saddle point which is not on the real axis, i.e. from a complex solution of the field equations. To see what this solution is, it is helpful to use the K representation of the wavefunction rather than the h_{ij} one.

The semiclassical approximation to the wave function $\Phi(k)$ is

$$\Phi(k) \approx N_k \sum_i A_i \exp - B_i, \tag{8.7}$$

where $k = \sigma K/9$, N_k is a normalization factor, B_i are the actions of classical solutions which are bounded by a three-sphere with the trace of the second fundamental form equal to K, and the prefactors A_i are given by the determinants of small fluctuations about the classical solutions. In the case under consideration, there is a unique classical solution for all values of K with action

$$\tilde{I}^k = -\frac{1}{3H^2}\left(1 - \frac{3k}{(9k^2+H^2)^{1/2}}\right]. \tag{8.8}$$

The wavefunction $\Psi(a)$ can then be obtained by an inverse Laplace transform of $\Phi(k)$,

$$\Psi(a) \approx \frac{N_k}{2\pi i}\int_\Gamma dk \exp(ka^3 - \tilde{I}^k), \tag{8.9}$$

where the contour Γ is parallel to the imaginary k-axis. The integral (8.9) can be evaluated by the method of steepest descent. If $Ha < 1$, there are two stationary phase points at real values of k which are equal in magnitude and are opposite in sign. They are the two values of k at which a three-sphere of radius a would fit into a four-sphere of radius

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

H^{-1}. The values of the exponent in (8.9) at these two stationary points are equal to $\tilde{I}_{\pm}(a)$ with the plus sign corresponding to $k < 0$ and the minus sign corresponding to $k > 0$, i.e. to filling in the three-sphere with more than half and with less than half the four-sphere, respectively.

There are complex stationary phase points of (8.9) but they give contributions which are smaller than those of the real stationary phase points described above. The steepest descent contour is shown in fig. 10.

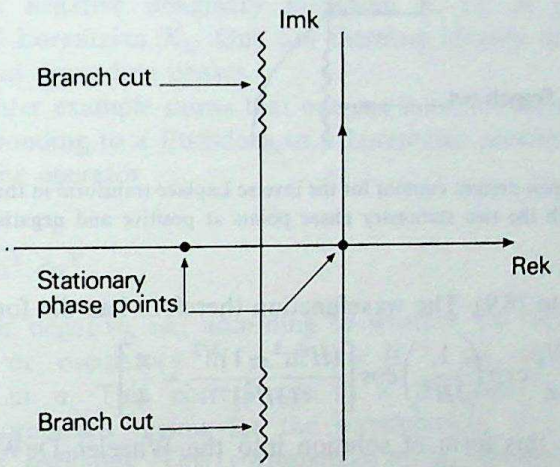


Fig. 10. The steepest descent contour for the inverse Laplace transform in the case $a < H^{-1}$. It passes through the stationary phase point at positive k but does not pass through that at negative k.

It passes through the stationary phase point at positive k but not through that at negative k. The value of the wavefunction is therefore

$$\Psi(a) \approx N_k \exp(-I_-(a)). \tag{8.10}$$

This agrees with (8.6).

If $Ha > 1$ there are no stationary phase points on the real k-axis because one cannot fit a three-sphere of radius $a > H^{-1}$ into a four-sphere of radius H^{-1}. There are however stationary phase points at

$$k = \pm \frac{i}{3} H \left(1 - \frac{1}{H^2 a^2}\right)^{1/2}. \tag{8.11}$$

It is possible to distort the contour Γ to pass through both of them (see fig. 11). There are other stationary phase points but they give a smaller

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

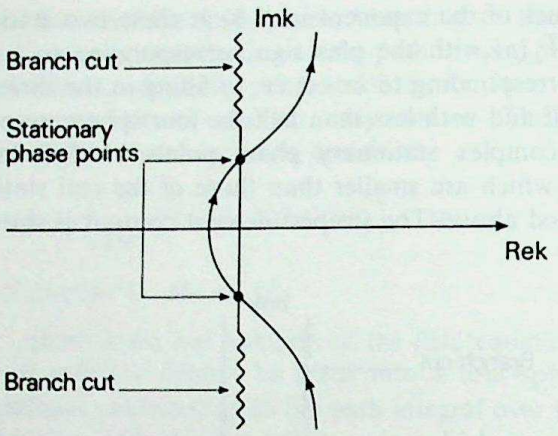


Fig. 11. The steepest descent contour for the inverse Laplace transform in the case $a > H^{-1}$. It passes through the two stationary phase points at positive and negative imaginary k.

contribution to (8.9). The wavefunction therefore has the form

$$\Psi(a) \approx \frac{N_k}{\pi} \exp\left(\frac{1}{3H^2}\right) \cos\left[\frac{(H^2a^2-1)^{3/2}}{(3H^2)} - \frac{\pi}{4}\right]. \tag{8.12}$$

One can put this form of solution into the Wheeler–DeWitt equation and solve for the amplitude prefactor A. One finds

$$A \approx \frac{1}{a^{1+p/2}}, \quad a \gg H^{-1}. \tag{8.13}$$

The inner product that makes the Wheeler–DeWitt operator Hermitian is

$$\int \Psi(a)\Psi^*(a)a^p \, da. \tag{8.14}$$

The wavefunction is therefore normalizable.

In order to interpret this wavefunction it is helpful to consider the classical Lorentzian solution with a Λ term. This is De Sitter space,

$$ds^2 = \sigma^2(-dt^2 + a^2 \, d\Omega_3^2), \tag{8.15}$$

where $a = H^{-1} \cos Ht$. It represents a space which collapses from infinite radius down to a minimum radius of H^{-1} at $t = 0$ and then re-expands exponentially. The wavefunction (8.6) and (8.12) is the quantum version of this. For a less than the classical minimum radius H^{-1} the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

wavefunction is exponentially small. This is what one would expect for a radius that was classically forbidden. For a greater than the minimum radius the wavefunction oscillates with constant amplitude. This means that any values of a greater than H^{-1} occur with roughly equal probability. In fact, one can write the cosine in (8.12) in the form

$$\exp(iHa^3/3)+\exp(-iHa^3/3). \tag{8.16}$$

The two terms in (8.16) come from the two stationary phase points at positive and negative imaginary Euclidean K, i.e. at positive and negative real Lorentzian K_L. One can therefore identify them with the collapsing and expanding phases.

The De Sitter example shows that one can interpret the wavefunction Ψ as corresponding to a Euclidean or a Lorentzian geometry according to whether the operator

$$K^2 = \frac{\partial}{\partial a^3}\,\frac{\partial}{\partial a^3}$$

is positive or negative, i.e., according to whether the wavefunction is exponential or oscillatory in a. If $a < H^{-1}$, the wavefunction is exponential in a. This corresponds to a Euclidean or classically forbidden Lorentzian regime for the wavefunction. If $a > H^{-1}$, the wavefunction oscillates. This corresponds to an allowed Lorentzian regime. Thus the wavefunction provides its own interpretation.

8.2. Our Universe?

The De Sitter example was useful because it showed how one could solve the Wheeler–DeWitt equation and apply the boundary conditions in a simple case. It also illustrated how one could interpret the wavefunction over three-geometries in terms of four-geometries in the classical limit. However, if there are two facts about our universe of which we are reasonably certain, one is that it is not exponentially expanding and the other is that it contains matter. In order to try to get a model which resembles our universe more closely, Hartle and I (Hartle and Hawking 1983) considered the next-simplest model, a homogeneous, isotropic universe with a conformally invariant scalar field ϕ which was constant on the three-sphere sections. The Wheeler–DeWitt equation takes the form

$$\frac{1}{2}\left[\frac{1}{a^p}\,\frac{\partial}{\partial a}\,a^p\,\frac{\partial}{\partial a} - a^2 + \lambda a^4 - \frac{\partial^2}{\partial \chi^2} + \chi^2\right]\psi[a,\chi] = 0, \tag{8.17}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where $\chi = (\pi a/\sigma)\phi\sqrt{2}$. The advantage of using a conformally invariant scalar field ϕ is that the Wheeler–DeWitt equation separates, i.e. there are solutions of the form

$$\psi(a, \chi) = C(a)f(\chi). \tag{8.18}$$

The function f obeys the harmonic oscillator equation

$$\left(-\frac{d^2}{d\chi^2} + \chi^2\right) f = Ef. \tag{8.19}$$

It is therefore natural to take f to be the harmonic oscillator wavefunction with eigenvalue $E = n + \frac{1}{2}$. The equation for the corresponding function C_n then is the Schroedinger equation for zero angular momentum in $p+1$ dimensions with a potential

$$V(a) = a^2 - \lambda a^4 \tag{8.20}$$

(see fig. 12).

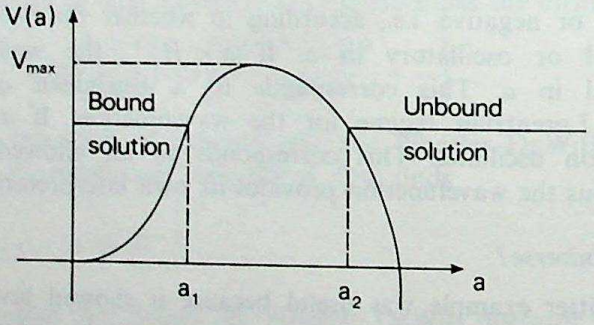


Fig. 12. The effective potential for the Wheeler–DeWitt equation for a minisuperspace model with a conformally invariant scalar field. If the separation constant E is less than the maximum of the potential, there are classically bound solutions with $a < a_1$ and unbound solutions with $a > a_2 > a_1$. Quantum mechanically, there is tunnelling between the bound and unbound solutions.

If the separation constant E is less than the maximum value V_{max} of $V(a)$, there are classically two kinds of allowed solutions: bound solutions in which a varies from zero up to some maximum value a_1, and unbound solutions in which a is always bigger than some minimum value $a_2 > a_1$. Quantum mechanically there will be tunneling between the bound and unbound solutions. One can therefore get a wavefunction which corresponds in the classical limit to a universe which expands

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

from $a = 0$ up to a maximum value $a = a_1$, and then recollapses to $a = 0$, but which has a small probability of tunneling through the potential barrier to a De Sitter-like state of indefinite expansion. In order for such a state to be a model of our universe Λ would have to be less than 10^{-120}, and n would have to be greater than 10^{120}. The tummeling probability would be of the order of $e^{-10^{120}}$.

One can ask which is the quantum state of the universe which is picked out by the boundary conditions of performing a path integral over compact metrics as described in section 6. It turns out that this is the state with $n = 0$. In this state the matter wavefunction is just

$$f(x) = e^{-x^2/2}. \tag{8.21}$$

and the gravitational wavefunction is just that of a De Sitter space as described in section 8.1. The quantum state determined by the boundary conditions of section 7 does not therefore resemble the universe that we live in. For this reason, Hartle and I proposed that this quantum state should be regarded as the "ground state" or state of minimum excitation of the universe but that the universe we live in is a linear superposition of the other "excited state" solutions with higher values of n. This meant that we lost any ability to predict the quantum state of the universe because we had no means of determining the complex coefficients of the excited states in the superposition.

The reason that the introduction of the scalar field ϕ does not change the gravitational wavefunction of the state determined by the boundary conditions of section 7 is that ϕ is conformally invariant. It therefore does not couple to the only gravitation degree of freedom in the model, which is a conformal degree of freedom. In order to get a wavefunction which is more like our universe, it seems necessary to put in matter fields which are not conformally invariant. I therefore considered the simplest example, which is a massive scalar field ϕ minimally coupled to the gravitational field. I got the idea for this from a talk given by Linde at the Shelter Island conference in June (Linde 1983). The Wheeler–DeWitt equation is

$$\frac{1}{2}\left[\frac{1}{a^p}\frac{\partial}{\partial a}a^p\frac{\partial}{\partial a} - a^2 + \lambda a^4 - \frac{1}{a^2}\frac{\partial^2}{\partial\phi^2}\right. \tag{8.22}$$

$$\left. + a^4 m^2 \phi^2\right]\Psi[a,\phi] = 0, \tag{8.23}$$

where m is the mass of the scalar field. One can regard this as a wave equation in the (a,ϕ) plane. The causal structure of the Wheeler–

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

DeWitt equation is seen more easily by introducing new coordinates,

$$x = a\sinh\phi, \qquad y = a\cosh\phi. \tag{8.24}$$

In these coordinates, the second derivatives take the form $\partial^2/\partial y^2 - \partial^2/\partial x^2$; thus, the characteristics are lines at 45° in the (x, y) plane. The region $a > 0$ is mapped into the interior of the future light cone of the origin (see fig. 13).

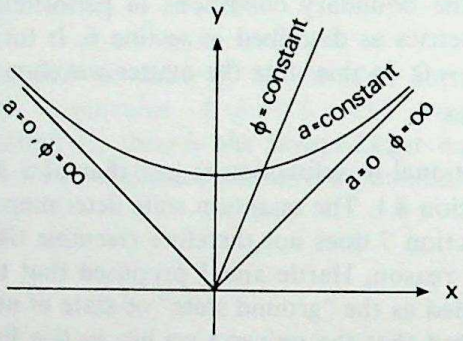


Fig. 13. The causal structure of the Wheeler–DeWitt equation can be seen in coordinates (x, y) in which the second derivatives take the form $\partial^2/\partial y^2 - \partial^2/\partial x^2$. The region $a > 0$ is mapped into the interior of the future light cone of the origin.

The surfaces of constant ϕ are straight lines through the origin and the surfaces of constant a are spacelike hyperbolae within the light cone.

If one knew the solution on the light cone of the origin, one could solve the wave equation by standard techniques. The light cone is the surface $a = 0$, $\phi = \pm\infty$. It is therefore rather hard to apply the boundary conditions of section 7 there, but what I shall do is use the semi-classical approximation to the path integral to estimate the form of the solution on lines of constant large positive or negative ϕ.

In order to apply the semi-classical approximation, one needs to know the solutions to the classical field equations. One can express the metric in the form

$$ds^2 = \sigma^2(d\tau^2 + a^2(\tau)d\Omega_3^2). \tag{8.25}$$

One is interested in metrics which are compact. Thus a will be zero at some value of τ which can be chosen to be zero. At $\tau = 0$, $d\phi/d\tau$ must be zero, one can then integrate the equations for $a(\tau)$ and $\phi(\tau)$ with the initial conditions

$$a = 0, \quad \frac{da}{d\tau} = 1, \quad \phi = \phi_0, \quad \frac{d\phi}{d\tau} = 0 \quad \text{at} \quad \tau = 0.$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

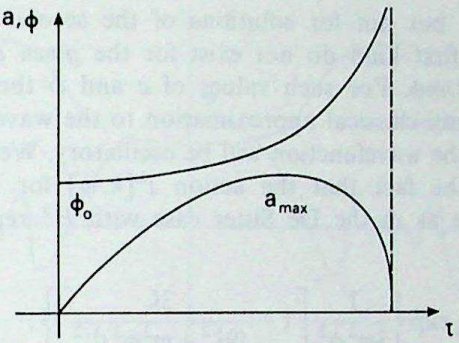


Fig. 14. Graph of the radius a and the scalar field ϕ for a solution of the Euclidean field equations.

As τ increases, $\phi(\tau)$ and $d\phi/d\tau$ will monotonically increase and $da/d\tau$ will monotonically decrease (see fig. 14).

What happens is that the positive energy density of the $m^2\phi^2(\tau)$ term in the action behaves like a positive cosmological constant and causes the four-geometry to have positive curvature. The three-sphere radius $a(\tau)$ rises to a maximum a_{max} which is a monotonically decreasing function of ϕ_0. The radius $a(\tau)$ then decreases and becomes zero at a singularity of the four-geometry. If $\phi_0 > 1$ then $\phi(\tau)$ does not increase much by the time $a(\tau)$ reaches a_{max}, and $a_{max} \approx 1/m\phi_0$ (see fig. 15).

In order to estimate the wavefuntion $\Psi[a, \phi]$, one looks for a solution $(a(\tau), \phi(\tau))$ which matches the given values of a and ϕ at some value $\tau = \tau_0$. There are two possible kinds of solutions: those for which $a(\tau)$ does not equal a_{max} in the range $0 \leqslant \tau \leqslant \tau_0$, and those for which it does. By an argument similar to the one in the De Sitter case, the semi-classical approximation to the wavefunction is given by e^{-I} for solutions

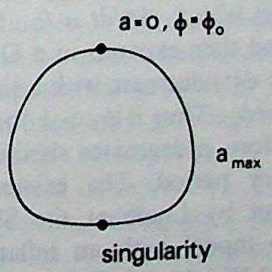


Fig. 15. A Euclidean solution of the field equations for a massive scalar field. The three-sphere radius a starts at zero at a regular point at the top of the diagram. It expands to a maximum radius a_{max} and then contracts to a singularity at zero radius.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

of the first kind but not for solutions of the second kind. However, solutions of the first kind do not exist for the *given* a and ϕ if both $\phi > 1$ and $a > 1/m\phi$. For such values of a and ϕ the solution which determines the semi-classical approximation to the wavefunction will be complex and so the wavefunction will be oscillatory. We can estimate its form by using the fact that the action $I^k[k, \phi]$ for $k > 0$, $\phi > 1$ is alsmost the same as in the De Sitter case with H^2 replaced by $m^2\phi^2$. Thus

$$\phi(k, \phi) \approx N_k \exp\left\{\frac{1}{3m^2\phi^2}\left[1 - \frac{3k}{(9k^2 + m^2\phi^2)^{1/2}}\right]\right\}. \tag{8.26}$$

One can do the inverse Laplace transform as before to find

$$\Psi(a, \phi) \approx N_k \exp\left(\frac{1}{3m^2\phi^2}\right) \cos(m\phi a^3),$$

this represents a De Sitter-type exponential expansion.

One can take this behaviour for large positive or negative ϕ as the boundary condition for the Wheeler–DeWitt equation near the light cone in the (x, y) plane. One can then solve the Wheeler–DeWitt equation as a wave equation with these boundary conditions to determine the behaviour of the wavefunction at other values of a and ϕ. Because the wavefunction is rapidly oscillating, one can use the WKB method. One writes the wavefunction in the form

$$\Psi(a, \phi) = C(a, \phi)\cos S(a, \phi), \tag{8.27}$$

where $S(a, \phi)$ is a rapidly varying phase and $C(a, \phi)$ is a slowly varying amplitude. The trajectories of ∇S in the (a, ϕ) plane correspond to classical Lorentzian solutions $(a(t), \phi(t))$ of the field equations. The semi-classical approximation for large ϕ corresponds to a classical solution which starts with $\phi(t) = \phi_1$, $\mathrm{d}\phi/\mathrm{d}t = 0$, at a minimum radius $a(t) = 1/m\phi_1$ at $t = 0$, and then expands in a De Sitter-like manner. The mass term tends to make $\phi(t)$ decrease with a timescale $3\phi_1/m$. However, the expansion rate $H = m\phi_1$. Thus if $\phi_1 > 1$ the universe will expand by a factor of order $\mathrm{e}^{3\phi_1^2}$ before ϕ decreases significantly. For large ϕ_1 this gives a long inflationary period. The existence of such a classical solution was pointed out by Linde at the Shelter Island conference. Another minisuperspace model with an inflationary period has been constructed by Moss and Wright (1983).

Eventually $\phi(t)$ decreases to zero and starts oscillating in time. The universe goes over to a matter-dominated phase with $a(t) \propto t^{2/3}$. The

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

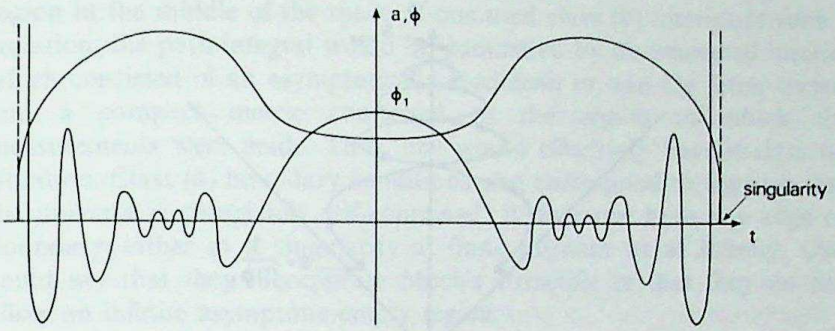


Fig. 16. A graph of the radius a and the scalar field ϕ for time symmetric solutions of the Lorentzian field equations. The radius expands from a minimum to a maximum and then, in general, collapses to a singularity.

universe will expand to a maximum radius of order $(1/m\phi_1)\exp(6\phi_1^2)$ and then recollapse. In general, it will recollapse to a singularity (see fig. 16). The solution is time symmetric about $t = 0$ but not in general time symmetry about $t = t_1$, the time of maximum expansion. However, for discrete values of the scalar field ϕ_1 the field $\phi(t)$ is zero at $t = t_1$. In these cases, the solution will be time symmetric about $t = t_1$ but with $\phi(t)$ replaced by $-\phi(t_1 - t)$. These solutions will therefore oscillate indefinitely. There are also periodic solutions which are not time symmetric about the minimum radius but which are symmetric about each maximum radius. The orbits of ∇S in the (a, ϕ) plane correspond to these periodic solutions. One thus obtains a wavefunction $\Psi(a, \phi)$ which can be interpreted as a superposition of wavefunctions corresponding to oscillating universes in the classical limit (see fig. 17).

The wavefunction $\Psi(a, \phi)$ will grow exponentially in a region near the y-axis in which $a^2m^2\phi^2 < 1$. This part of the wavefunction can be interpreted as corresponding to a flat Euclidean four-geometry. If one wished the wavefunction to be normalizable, one could introduce a tiny λ term to cut off the exponential growth at very large values of a. Superimposed on this exponentially growing wavefunction will be an oscillating component which corresponds to the periodic solutions described above. As was stated in section 8.1, the oscillating component should be interpreted as corresponding to a Lorentzian geometry and the exponential component as corresponding to a Euclidean geometry. We live in a Lorentzian geometry and therefore we are interested really only in the oscillatory part.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

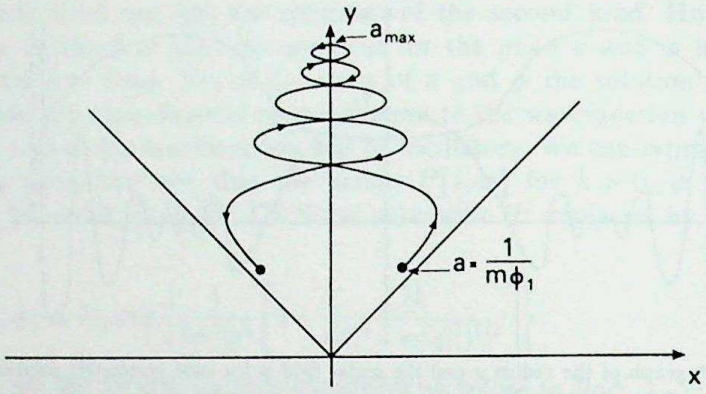


Fig. 17. A WKB orbit of ∇S in the (x, y) plane. It corresponds to a periodic Lorentzian solution of the field equations with a long period of exponential expansion.

9. Conclusion

The hot Big Bang model seems to provide a good description of the universe, at least, since the time of light-element formation. However, it has a number of features, such as the large-scale homogeneity and isotropy, which are not explained but are simply assumed as initial conditions. Some of these features can be accounted for in the Inflationary Model which postulates that there was a period of exponential expansion at the GUT era. However, it was still necessary to assume that the universe started off in thermal equilibrium at very high temperature and that it was sufficiently smooth, isotropic and flat that it expanded until the temperature fell to the Grand Unified Energy of around 10^{15}GeV. Any discussion of the initial conditions must involve quantum gravity and a specification of the quantum state of the universe. The quantum state is determined by the class of metrics and field configurations that are summed over in the path integral. There seem to be only two natural classes of Euclidean, i.e. positive definite, metrics:

(a) Compact metrics;

(b) Metrics that are asymptotic to a non-compact metric of maximal symmetry such as flat Euclidean space or Euclidean anti-De Sitter space.

Class (b) is appropriate to scattering problems in which one sends in particles from infinity and measures what comes back to infinity. However, it does not seem to be suitable for cosmological problems where one is concerned with measurements, not at infinity, but in a finite

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

region in the middle of the space. If one used class (b) metrics in such a situation, the path integral would be dominated by disconnected metrics which consisted of an asymptotically Euclidean or anti-De Sitter metric and a compact metric connected to the region in which the measurements were made. Thus, one would effectively have a class (a) situation. Class (a) boundary conditions also correspond to the idea that the universe is completely self-contained: it does not have any edge or boundary, either at a singularity at finite distance or at infinity. One could say that they incorporate Mach's Principle in that they do not allow an infinite asymptotic empty region.

The ultimate test of any proposed boundary conditions for the universe, however, is not that they are aesthetically appealing but that they predict a universe which corresponds to what we observe. In section 8.2 I showed that in a simple minisuperspace model of a homogeneous isotropic universe with a massive scalar field the boundary conditions led to a wavefunction which could be interpreted as a superposition of states which corresponded to oscillating universes in the classical limit. One might object that the wavefunction corresponded to a whole family of different classical solutions rather than being sharply peaked around one unique solution. One could ask: which solution do we observe? I think however that one can resolve this apparent paradox in the following way: imagine a quantum state that was sharply peaked around a particular classical solution of the field equations. The wavefunction for this state would also describe any intelligent beings who existed in this universe and who measured its properties. Now imagine another quantum state that was peaked around a different classical solution. Any measurements made by intelligent beings in this quantum state would correspond to the properties of the second solution. One could now consider a quantum state that was a superposition of the two states peaked around different classical solutions. Because of the linearity of quantum mechanics, there would be no interference: measurements made by the intelligent beings in the first universe would correspond to the properties of the first classical solution and measurements made in the second universe would correspond to the second solution. What the specification of the boundary conditions, or equivalently, the quantum state of the universe does, is it determines the relative probability of universes corresponding to different classical solutions. In the case of the model considered in section 8.2, this probability is proportional to something of the order of $\exp(m^{-2}\phi_1^{-2})$. This is fairly flat for $\phi_1 > m^{-1}$ (there may be polynomial factors of ϕ_1). We must live in a solution which expands to a radius of at least 10^{28} cm.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Given this, the probability is then high that the universe will expand to a radius much greater than this, i.e. that we live in a universe with Ω very near one.

The model of section 8.2 has a long period of exponential expansion. One would expect that this would reduce any inhomogeneties or anisotropics if one put in these extra degrees of freedom. Thus it would have all the features of the Inflationary Model described in section 4, with the additional advantage that it arose from a natural choice of boundary conditions for the universe. The classical solutions that correspond to the wavefunction are periodic. Thus, if entropy or disorder increases during the expansion phase, it must decrease during the contraction phase. In fact this is a general feature of any wavefunction determined by a Euclidean boundary condition like (a) or (b) above. The wavefunction $\Phi(\tilde{h}_{ij}, K)$ must be real for real Euclidean K. This means that $\Phi(\tilde{h}_{ij}, K_L) = \Phi^*(\tilde{h}_{ij}, -K_L)$ for real Lorentzian K_L. One can interpret this as implying that the state is time-symmetric in the sense that the wavefunction in the contracting phase is the complex conjugate of the wavefunction in the expanding phase. Despite this time-symmetry it is possible that there could be a well-defined arrow of time as measured by local entropy increase during the expansion phase which reversed during the contraction. Whether or not this is the case and whether or not the arrow agrees with the direction of time defined by the expansion remains to be shown.

There are many objections that can be raised against the minisuperspace model of section 8.2, such as that an infinite number of degrees of freedom have been neglected and all questions of finiteness or renormalizability have been ignored. In my opinion the significance of the model is not necessarily that it describes the actual universe but that it shows that the boundary conditions (a) above can give rise to a quantum state that corresponds to the universe that we live in. This and the fact that there does not seem to be any natural alternative set of boundary conditions encourages me to believe that the boundary conditions (a) are correct. All that we would then need is a complete consistent theory of quantum gravity and the other interactions, and we would be able to predict everything, at least in principle.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

References

Albrecht, A., and P.J. Steinhardt, 1982, Phys. Rev. Lett. **48**, 120.
Bardeen, J.M., P.J. Steinhardt and M.S. Turner, 1982, preprint URS-2020T (Univ. of Pennsylvania).
Coleman, S., 1977, Phys. Rev. **D15**, 2929.
Coleman, S., and E. Weinberg, 1973, Phys. Rev. **D7**, 188.
Eguchi, T., P.B. Gilkey and A.J. Hansen, 1980, Phys. Rev. **66**, 214.
Gibbons, G.W., and S.W. Hawking, 1977, Phys. Rev. **D15**, 2752.
Gibbons, G.W., S.W. Hawking and M.J. Perry, 1978, Nucl. Phys. **B138**, 141.
Gibbons, G.W., S.W. Hawking, G.W. Horowitz and M.J. Perry, 1983, Commun. Math. Phys. **88**, 295.
Guth, A.H., 1981, Phys. Rev. **D23**, 347.
Guth,A.H., and So-Young Pi, 1982, Phys. Rev. Lett. **49**, 1110.
Guth, A.H., and E. Weinberg, 1981, Phys. Rev. **D23**, 876.
Harrison, E.R., 1970, Phys. Rev. **D1**, 2726.
Hartle, J.B., and S.W. Hawking, 1983, Phys. Rev. **D28**, 2960.
Hawking, S.W., 1979a, The Path-Integral Approach to Quantum Gravity, in: General Relativity: An Einstein Centenary Survey, eds. S.W. Hawking and W. Israel (Cambridge Univ. Press).
Hawking, S.W., 1979b, Euclidean Quantum Gravity, in: Recent Developments in Gravitation, Cargese Lectures, eds. M. Levy and S. Deser (Plenum, New York).
Hawking, S.W., 1982, in: Astrophysical Cosmology, Pontificiae Academiae Scientiarum Varia 48, 563.
Hawking, S.W., 1983a, Phil. Trans. Roy. Soc. **A310**, 303–310.
Hawking, S.W., 1983b, Phys. Lett. **134B**, 403.
Hawking, S.W., and G.F.R. Ellis, 1973, The Large Scale Structure of Space-Time (Cambridge Univ. Press).
Hawking, S.W., and I.G. Moss, 1982, Phys. Lett. **110B**, 35.
Hawking, S.W. and I.G. Moss, 1983, Nucl. Phys. **B224**, 180–192.
Hawking, S.W., I.G. Moss and J.M. Stewart, 1982, Phys. Rev. **D10**, 2681.
Linde, A.D., 1982, Phys. Lett. **108B**, 389.
Linde, A.D., 1983, Phys. Lett. **129B**, 177.
Moss, I.G., and W.A. Wright, 1983, The Wave Function of the Inflationary Universe, preprint (Newcastle Univ.).
York, J.M., 1972, Phys. Rev. Lett. **28**, 1082.
Zel'dovich, Ya.B., 1972, Mon. Not. R. Ast. Soc. **160**, 1.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Nuclear Physics B239 (1984) 257–276
© North-Holland Publishing Company

THE QUANTUM STATE OF THE UNIVERSE

S.W. HAWKING

Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, England

Received 28 December 1983

The quantum state of the universe is determined by the specification of the class of metrics and matter field configurations that are summed over in the path integral. The only natural choice of this class seems to be compact euclidean (i.e. positive definite) metrics and matter fields that are regular on them. This choice incorporates the idea that the universe is completely self-contained and has no boundary or asymptotic region. I show that in a simple "minisuperspace" model this boundary condition leads to a wave function which can be interpreted as a superposition of quantum states which are peaked around a family of classical solutions of the field equations. These solutions are non-singular and represent oscillating universes with a long inflationary period. They could be a good description of the observed universe. I also show that the features of the minisuperspace model that give rise to such a wave function are also present in models that contain all the degrees of freedom of the gravitational and matter fields.

1. Introduction

The task of theoretical physics is to construct a mathematical model of the universe which agrees with all the observations made so far and which predicts the results of future observations. This model usually consists of two parts:

(1) Local laws which govern the physical fields in the model. In classical physics, these laws are normally expressed as differential equations which can be derived from an action I. In quantum physics the laws can be obtained from a path integral over all field configurations weighted with $\exp(iI)$.

(2) Boundary conditions which pick out one particular state from among the set of those allowed by the local laws. The classical state can be specified by boundary conditions for the differential equations and the quantum state can be determined by asymptotic conditions on the class C of field configurations that are summed over in the path integral.

A great deal of work has been done on the first part, particularly in the last twenty years. Although we do not yet have a complete, consistent field theory, we probably know most of the features that it should have. The outstanding remaining problem is to construct a quantum theory of gravity. The standard approach to quantizing general relativity by a perturbation expansion produces infinities which

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

are "unrenormalizable", that is, they cannot be removed by redefining the quantities in the action. There are three possible solutions to this problem:

(i) It may be possible to find a theory, such as $N = 8$ supergravity, in which all the infinities cancel to all orders in the perturbation expansion.

(ii) The infinities may simply be a result of the perturbation expansion which is known to break down at high energies. Gravitational effects may cut off the high energies and give rise to a finite theory.

(iii) It may be necessary to add higher derivative terms to the gravitational action. This is known to give a renormalizable theory but there are problems with negative energies, unitarity and runaway solutions.

It seems likely that the solution to the problems of quantum gravity lies with one of the three alternatives listed above. My personal opinion is that it is probably the second possibility: the path integral over the conformal factor seems to damp out the contributions of metrics with a lot of conformal curvature. However, whichever of the three solutions is the correct one, it is likely that we will know before too long.

By contrast with all the work on the local laws, very little has been done on the boundary conditions. Indeed, many people would claim that the boundary conditions are not part of physics but belong to metaphysics or religion. They would claim that nature had complete freedom to start the universe off any way it wanted. That may be so, but it could also have made it evolve in a completely arbitrary and random manner. Yet all the evidence is that it evolves in a regular way according to certain laws. It would therefore seem reasonable to suppose that there are also laws governing the boundary conditions. One might also argue that the observed universe is so complicated that it could not possibly have arisen from some simple boundary condition. I do not think that this is a valid objection: the laws of quantum electrodynamics are simple but they give rise to all the complexities of chemistry and biology. I shall show in a simplified model that a simple boundary condition for the universe can still give rise to very complicated behaviour. Indeed, a number of classically allowed behaviours occur with relative probabilities determined by the boundary conditions.

It might seem premature to speculate about the boundary conditions when we are not yet sure of the exact form of the local laws. However, we know that the large-scale structure of the universe is determined by gravity and we are fairly sure that gravity is described by general relativity or some variant, such as supergravity or higher derivative theories. We also believe that quantum theory can be formulated in terms of path integrals. The problem of boundary conditions then becomes one of specifying the class of spacetime metrics and matter field configurations that are summed over in the path integral. I shall adopt the point of view that the path integral should be evaluated in the euclidean regime, that is, over positive definite metrics and matter fields in these metrics. There are then only two natural choices for the class of metrics:

(a) compact metrics and matter fields that are regular on them;

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

(b) non-compact metrics that are asymptotic to metrics of maximal symmetry, i.e. to flat euclidean space or to euclidean anti-de Sitter space, and matter fields that are asymptotically zero.

Boundary conditions of type (b) give rise to the usual vacuum state that one uses in scattering problems in which one sends particles in from infinity and measures what comes back out to infinity. However, I shall show in sect. 3 that they are not suitable as boundary conditions for the whole universe. This leaves only boundary conditions of type (a). They incorporate the idea that the universe is completely self-contained and has no boundaries, either at singularities at finite distance or at infinity. One could paraphrase them as: "the boundary conditions of the universe are that it has no boundary".

Boundary conditions of type (a) were proposed in [1] but at that time I did not know how to calculate their consequences. The necessary formalism was developed in collaboration with Hartle in [2] and is summarised in sect. 2. The formalism is applied in sect. 4 to a simple "minisuperspace" toy model of a homogeneous isotropic universe containing a massive scalar field. It is shown that the boundary conditions (a) give rise to a wave function that can be interpreted as a superposition of quantum states peaked around different classical solutions representing oscillating universes. One might object that the quantum state should be peaked around only a single classical solution: after all, we observe only one universe. The answer is that if one had a quantum state that was peaked around a particular classical solution, that quantum state would describe not only the universe but also any observers who measured its properties. Suppose one now had a different quantum state that was peaked around a different classical solution of the field equations. That quantum state would also contain observers who would measure that properties of the second solution. Then consider a quantum state that was the sum of the two previous states. Because of the linearity of quantum mechanics, there would be no interference; observers in the first state would measure the properties of the first solution and observers in the second state would measure those of the second. The classical solutions that correspond to the wave function of the model can have a long period of exponential expansion and could be a reasonable description of the universe that we live in. They would have most of the desirable features of the inflationary model [3–7] and the advantage of having been derived from a definite proposal for the boundary conditions of the universe.

One could easily object to the minisuperspace toy model on the grounds that it truncated the infinite number of degrees of freedom of the gravitational and matter fields down to a finite number and that it ignored the problems of the divergences of quantum gravity. However, I show in sect. 5 that the features of the minisuperspace model that lead to a wave function of the right form would also be present in a model that did not restrict the degrees of freedom provided that it contains fields of non-zero rest-mass. The quantum corrections to the gravitational action will produce terms of the form R^2 and $C_{\mu\nu\rho\lambda}\, C^{\mu\nu\rho\lambda}$. These will be finite in the case of

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

possibilities (i) and (ii) and infinite but renormalized in the case of possibility (iii). Such terms can act like massive scalar or spin-2 fields. It will be shown in a future paper [8] that the R^2 term at least leads to a minisuperspace model which is very similar to that with a massive scalar field.

These results and the fact that there does not seem to be any other natural candidate leads me, at least, to believe that the quantum state of the universe is determined by a path integral over compact positive definite metrics. If this is correct, the second part of the problem of constructing a mathematical model of the universe will have been solved. It would then only remain to determine the exact form of the local laws and we should, in principle, be able to predict the probability of anything in the universe. In practice of course, the equations will be much too complicated to solve in any but very simple situations.

2. Quantum gravity

In certain special spacetime metrics, such as those of the Schwarzschild or de Sitter solutions, it is possible to define a new time-coordinate $\tau = it$ and to change the metric from a lorentzian signature $(-+++)$ to a euclidean or positive definite signature $(++++)$. However, this is not possible in general: a general lorentzian metric will not have a section in the complexified spacetime manifold on which the metric is real and positive definite. Similarly, a general positive definite metric will not have a section on which the metric is real and lorentzian. In this respect, the spacetime metric behaves like other fields in ordinary quantum field theory in flat space: a Yang–Mills field which is real in Minkowski space will not, in general, be real when analytically continued to euclidean space and a field which is real in euclidean space will not be real when analytically continued to Minkowski space. The point about the euclidean approach to quantum field theory in flat space is not that individual configurations in euclidean space correspond with configurations in Minkowski space, but that the path integral over all real field configurations in euclidean space is equivalent, in the sense of contour integration, to the path integral over all real field configurations in Minkowski space. If this works for other quantum fields, it seems reasonable to suppose that it works for gravity also. I shall therefore assume that the quantum theory of the universe can be derived from a path integral over euclidean, i.e. positive definite, metrics and matter fields which are regular on the manifolds defined by these metrics. I shall show that the Wheeler–DeWitt equation, the analogue of the Schrödinger equation, is the same whether it is derived from a path integral over euclidean or lorentzian metrics but that the wave function of the universe itself determines whether it corresponds to a euclidean or a lorentzian geometry in the classical limit.

As in refs. [2, 9], my starting point is the assumption that the probability for a 4-metric $g_{\mu\nu}$ and a matter field configuration ϕ is proportional to

$$\exp\left(-\tilde{I}[g_{\mu\nu}, \phi]\right), \tag{2.1}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where $\tilde{I}$ is the euclidean action

$$\tilde{I}[g_{\mu\nu}, \phi] = \frac{m_p^2}{16\pi}\left(-\int_{\partial M} 2Kh^{1/2}\, d^3x - \int_M \left(R - 2\Lambda + \frac{16\pi}{m_p^2} L(g_{\mu\nu}, \phi)\right) g^{1/2}\, d^4x\right), \tag{2.2}$$

where h_{ij} is the 3-metric on the boundary ∂M and K is the trace of the second fundamental form of the boundary. The surface term in the action is necessary because the curvature scalar R contains second derivatives of the metric [10–12]. The physics of the universe is governed by probabilities of the form (2.1) for all 4-metrics $g_{\mu\nu}$ and matter field configurations belonging to a certain class C. The specification of the class C determines the quantum state of the universe. This will be considered further in sect. 3.

In practice, one is normally interested in the probability, not of the entire 4-metric, but of a more restricted set of observables. Such a probability can be derived from the basic probability (2.1) by integrating over the unobserved quantities. In cosmology, one is concerned with observables, not at infinity, but in some finite region in the interior of the 4-geometry. A particularly important case is the probability $P[h_{ij}, \phi]$ of finding a closed compact 3-submanifold S which divides the 4-manifold M into two parts $M_\pm$ and on which the induced 3-metric is h_{ij} and the matter field configuration is ϕ:

$$P[h_{ij}, \phi] = \int d[g_{\mu\nu}]\, d[\phi] \exp\left(-\tilde{I}[g_{\mu\nu}, \phi]\right), \tag{2.3}$$

where the integral is over all 4-metrics and matter field configurations belonging to the class C which contain a 3-submanifold S on which the induced 3-metric is h_{ij} and the matter field configuration is ϕ. This probability can be factorized into the product of two amplitudes or wave functions $\Psi_\pm[h_{ij}, \phi]$, $P[h_{ij}, \phi] = \Psi_+[h_{ij}, \phi]\Psi_-[h_{ij}, \phi]$, where

$$\Psi_\pm[h_{ij}, \phi] = \int_{C_\pm} d[g_{\mu\nu}]\, d[\phi] \exp\left(-\tilde{I}[g_{\mu\nu}, \phi]\right). \tag{2.4}$$

The path integral is over the classes $C_\pm$ of 4-metrics and matter field configurations on $M_\pm$ which agree with the given 3-metric h_{ij} and matter field configuration ϕ on S. For example, if the class C consisted of asymptotically euclidean metrics, then C_+ would consist of 4-metrics and matter field configurations on compact manifolds M_+ with boundary S and C_- would consist of metrics and fields on asymptotically euclidean manifolds M_- with inner boundary S. In this case, the two wave functions $\Psi_\pm$ will be different from each other but, if the class C consists of compact metrics, they will be the same.

One can regard $\Psi_\pm(h_{ij}, \phi]$ as the "wave functions of the universe" [2]. Note that they do not depend on time or, more generally, on the position of the 3-manifold S in the 4-manifold M. In general, there is no invariant way to specify this. In fact,

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the 3-metric h_{ij} usually determines the position of S in M or determines it up to a finite ambiguity. In what follows I shall drop the subscripts $\pm$ on the wave function Ψ.

The wave function $\Psi[h_{ij}, \phi]$ obeys a functional differential equation, the Wheeler–DeWitt equation, which is the analogue of the Schrödinger equation. This can be derived from the path integral definition (2.4) as follows. In the neighbourhood of the surface S, one can introduce a time-coordinate τ which is constant on S so that the metric takes the standard 3+1 form:

$$ds^2 = (N^2 - N_i N^i)\, d\tau^2 + 2N_i\, dx^i\, d\tau + h_{ij}\, dx^i\, dx^j. \tag{2.5}$$

The action then becomes

$$\tilde{I}[g_{\mu\nu}, \phi] = \frac{m_p^2}{16\pi} \int d^3x\, d\tau\, h^{1/2} N [K_{ij}K^{ij} - K^2 - {}^3R(h_{ij}) + 2\Lambda - 16\pi m_p^{-2} L(\phi)], \tag{2.6}$$

where K_{ij} is the second fundamental form

$$K_{ij} = \frac{1}{N}\left[\frac{1}{2}\frac{\partial h_{ij}}{\partial \tau} + N_{(i|j)}\right] \tag{2.7}$$

and a stroke and ${}^3R$ denote the covariant derivative and scalar curvature constructed from the 3-metric h_{ij}.

The functional integral defining the wave function contains an integral over N. By varying N at the surface one pushes it forward or backward in time. Since the wave function does not depend on time one must have

$$0 = \int d[g_{\mu\nu}]\, d[\phi] \left[\frac{\delta \tilde{I}}{\delta N}\right] \exp(-\tilde{I}[g_{\mu\nu}, \phi]). \tag{2.8}$$

Classically the field equation $H \equiv \delta\tilde{I}/\delta N = 0$ is the hamiltonian constraint for general relativity. It is

$$H = -\frac{m_p^2 h^{1/2}}{8\pi}(K^{ij}K_{ij} - K^2 + {}^3R - 2\Lambda + 8\pi m_p^{-2} T_{nn}) = 0, \tag{2.9}$$

where T_{nn} is the euclidean stress energy tensor of the matter field projected in the direction normal to the surface. One can substitute the expression (2.9) for $\delta\tilde{I}/\delta N$ in (2.8). The second fundamental form K_{ij} can be expressed in terms of the functional operator $\delta/\delta h_{ij}$:

$$\frac{\delta \tilde{I}}{\delta h_{ij}} = \frac{m_p^2}{16\pi} h^{1/2}(K_{ij} - h_{ij}K). \tag{2.10}$$

This identity can be derived by varying the metric in the definition of the action (2.2). The time derivative of the variation of the metric can be removed by integrating by parts in the normal manner. This produces a surface term which is δh_{ij} times

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the right-hand side of (2.10). Similarly, the time derivatives of the matter fields ϕ can be expressed in terms of the operator $\delta/\delta\phi$. Using these substitutions in (2.8), one obtains the Wheeler-DeWitt equation

$$\left[-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}} - h^{1/2}\left({}^3R(h) - 2\Lambda + 8\pi m_{\mathrm{p}}^{-2} T_{\mathrm{nn}}\left(\frac{\delta}{\delta\phi}, \phi\right)\right)\right]\Psi[h_{ij}, \phi] = 0 , \tag{2.11}$$

where G_{ijkl} is the metric on superspace, the space of all 3-metrics h_{ij}:

$$G_{ijkl} = \tfrac{1}{2}h^{1/2}(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}) . \tag{2.12}$$

There is a question of the factor ordering in the functional differential equation or, equivalently, of the measure in the path integral. This will not be important in the situation that I shall consider. The Wheeler–DeWitt equation is the same as the equation one would have obtained had one started out with a path integral over lorentzian metrics. One can also consider variations within the 3-surface S generated by the "Shift" vector N_i:

$$0 = \int \frac{\delta\tilde{I}}{\delta N_i} \exp(-\tilde{I}[g_{\mu\nu}, \phi]) . \tag{2.13}$$

This equation implies that the wave function is invariant under diffeomorphisms, i.e. that Ψ is a functional of the 3-geometry and not of the particular 3-metric h_{ij}.

The metric $-G_{ijkl}$ on superspace, the space of 3-geometries, has a signature $(+-----)$. The Wheeler–DeWitt equation can therefore be thought of as a hyperbolic equation on superspace with $h^{1/2}$ as the time coordinate. I shall show that the semiclassical approximation to the path integral gives the boundary conditions for the Wheeler–DeWitt equation at small $h^{1/2}$. The Wheeler–DeWitt equation can then be solved to give the values of the wave function at larger values of $h^{1/2}$. In a lorentzian metric, $h^{1/2} = 0$ would be a singularity. However, this is not necessarily the case in a euclidean metric as can be seen by considering the example of a 4-sphere of radius R embedded in flat 5-dimensional space: a surface $|x^5| < R$ intersects the 4-sphere in a 3-sphere of non-zero radius. However, when $|x^5| = R$, the 3-sphere shrinks to zero radius but there is no singularity of the 4-geometry. Indeed, $h^{1/2}$ has to go to zero if the topology of the 3-surfaces is to change.

By its construction, the wave function $\Psi[h_{ij}, \phi]$ vanishes for 3-metrics h_{ij} which are not positive definite, i.e. for which the determinant h of the metric is negative. This means that h_{ij} is not a quantum observable with an unrestricted range and that in many cases it is more convenient to replace it by $\tilde{h}_{ij}$, the 3-metric up to a conformal factor, and $Km_{\mathrm{p}}^2/12\pi$, the momentum conjugate to $h^{1/2}$. One can then define a wave function Φ in this representation by

$$\Phi[\tilde{h}_{ij}, K, \phi] = \int \mathrm{d}[g_{\mu\nu}]\,\mathrm{d}[\phi] \exp(-I^k[g_{\mu\nu}, \phi]) , \tag{2.14}$$

where $I^k[g_{\mu\nu}, \phi]$ is the action appropriate to the situation in which K and $\tilde{h}_{ij}$ are

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

kept fixed on the boundary rather than h_{ij}. It differs from the action $\tilde{I}[g_{\mu\nu}, \phi]$ in that the surface term has a different coefficient, $\frac{2}{3}K$ rather than $2K$. The wave function $\Phi[\tilde{h}_{ij}, K, \phi]$ is the Laplace transform of $\Phi[h_{ij}, \phi]$:

$$\Phi[\tilde{h}_{ij}, K, \phi] = \int_0^\infty \mathrm{d}[h^{1/2}] \exp\left[-\frac{m_p^2}{12\pi} \int \mathrm{d}^3x\, h^{1/2} K \right] \Psi[h_{ij}, \phi]. \qquad (2.15)$$

Similarly, $\Psi[h_{ij}, \phi]$ is the inverse Laplace transform of $\Phi[\tilde{h}_{ij}, K, \phi]$.

$$\Psi[h_{ij}, \phi] = \int_\Gamma \mathrm{d}\left[\frac{m_p^2}{24\pi i} K \right] \exp\left[\frac{m_p^2}{12\pi} \int \mathrm{d}^3x\, h^{1/2} K \right] \Phi[\tilde{h}_{ij}, K, \phi], \qquad (2.16)$$

where for each point of S the contour Γ runs from $-i\infty$ to $+i\infty$ to the right of any singularities of $\Phi[\tilde{h}_{ij}, K, \phi]$ in the complex K plane. Providing that $\Phi[\tilde{h}_{ij}, K, \phi]$ does not diverge exponentially for large Re K, this choice of contour will ensure that $\Psi[h_{ij}, \phi] = 0$ for $h^{1/2} < 0$ because one can close the contour in the right half K plane.

The square of the trace of the second fundamental form K^2 can be expressed as an operator on the wave function Ψ:

$$K^2 = \frac{64\pi^2}{m_p^4} \left[h^{-1/2} h_{ij} \frac{\delta}{\delta h_{ij}} \right] h^{-1/2} h_{kl} \frac{\delta}{\delta h_{kl}}. \qquad (2.17)$$

If $K^2\Psi/\Psi$ is positive, one can interpret the wave function Ψ in terms of a euclidean 4-geometry in the classical limit. On the other hand, if $K^2\Psi/\Psi$ is negative, then the observable K is imaginary which corresponds to a lorentzian 4-geometry in the classical limit. The quantity $K^2\Psi/\Psi$ is positive or negative according to whether the wave function Ψ depends in an exponential or oscillatory manner on the scale $h^{1/2}$. Thus the form of the wave function determines whether it corresponds to a euclidean or lorentzian 4-geometry in the classical limit.

3. Boundary conditions

As was stated in the introduction, the boundary conditions on the class C of metrics that are summed over in the path integral determine the quantum state of the universe. There seem to be only two natural boundary conditions for positive definite metrics:

(a) compact metrics;

(b) non-compact metrics which are asymptotic to metrics of maximal symmetry, i.e. flat euclidean space or euclidean anti-de Sitter space.

Boundary conditions of type (b) define the usual vacuum state. In this state, the expectation values of most quantities are defined to be zero so the vacuum state is not of much interest as the quantum state of the universe. In particle scattering calculations, one starts with the vacuum state and one changes the state by creating particles by the action of field operators at infinity in the infinite past. One lets the particles interact and then annihilates the resultant particles by the action of other

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

field operators at future infinity. This gets one back to the vacuum state. If one supposed that the quantum state of the universe was some such particle scattering state, one would lose all ability to predict the state of the universe because one would have no idea of what was coming in. One would also expect that the matter in the universe would be concentrated in a certain region and would decrease to zero at large distances instead of the roughly homogeneous universe that we observe.

In particle scattering problems, one is interested in observables at infinity. One is therefore concerned only with metrics which are connected to infinity: any disconnected compact parts of the metric would not contribute to the scattering of particles from infinity. In cosmology, on the other hand, one is concerned with observables in a finite region in the middle of the space and it does not matter whether this region is connected to an infinite asymptotic region. For example, one might ask for the amplitudes $\Psi_{\pm}[h_{ij}, \phi]$ for a closed compact 3-manifold S with a 3-metric h_{ij} and a matter field configuration ϕ. As explained in sect. 2, $\Psi_{\pm}$ is given by a path integral over all metrics belonging to classes $C_{\pm}$ which are bounded by S with the given 3-metric and matter field configuration. If the class C that determines the quantum state consists of metrics of type (b), then C_+ would consist of compact metrics and C_- would contain two different kinds of metrics:

(i) connected asymptotically euclidean or anti-de Sitter metrics which had an inner boundary at S;

(ii) disconnected metrics which consisted of a compact part with boundary at S and an asymptotically euclidean or anti-de Sitter part without any inner boundary.

One cannot exclude disconnected metrics from the path integral because they can be approximated by connected metrics in which the different parts were joined by thin tubes. The tubes could be chosen to have negligible action. Similarly, topologically non-trivial metrics cannot be excluded because they can be approximated by topologically trivial metrics.

One would expect the dominant contribution to the path integral for the wave function to come from metrics which were near to solutions of the field equations with the appropriate boundary conditions: In the case of metrics of kind (i), it is shown in the appendix however that such solutions cannot give the dominant contribution to the wave function. This is borne out by calculations in specific examples. Thus, the dominant contribution to the wave function would come from metrics of kind (ii). These in general have a smaller action and hence give a bigger contribution than metrics of kind (i). One sees therefore that even if one adopted boundary conditions of type (b), the result would be almost the same as if one had adopted type (a), as far as the amplitude for finding a compact 3-submanifold S was concerned. Similar results would apply to other observations in a finite region. It would therefore seem more natural to suppose that the boundary conditions of the universe were of type (a), i.e. that the quantum state of the universe was defined by a path integral over compact positive definite metrics. In this case, the two amplitudes $\Psi_{\pm}$ are equal. I shall therefore drop the subscript $\pm$.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

It should be emphasized that this is a *proposal* for the quantum state of the universe. One cannot derive it from some other principle but merely show that it is a natural choice. The ultimate test, however, is not whether it is aesthetically appealing but whether it enables one to make predictions that agree with observations. I shall endeavour to do this in the next section for a simple model.

4. Minisuperspace

In order to investigate whether the boundary conditions proposed in the last section correspond to the quantum state of the universe that we live in, one would like to calculate the wave function $\Psi[h_{ij}, \phi]$ and see if it can be interpreted as predicting what we observe. This is equivalent to finding a solution of the Wheeler–DeWitt equation obeying certain boundary conditions. The Wheeler–DeWitt equation is a second order functional differential equation on an infinite dimensional manifold, superspace, the space of all three metrics and matter field configurations. We do not know how to solve such an equation but we can hope to get some idea of the nature of the solution by considering the equation restricted to a finite dimensional submanifold called, minisuperspace. In other words, one restricts the infinite number of degrees of freedom of the gravitational and matter fields down to a finite number. The Wheeler–DeWitt equation then becomes a wave equation on a finite dimensional manifold and can be solved by standard techniques. The boundary conditions for the Wheeler–DeWitt equation can be obtained from the semiclassical approximation to the path integral, i.e. by expressing the wave function in the form

$$\Psi[h_{ij}, \phi] = N_0 \sum_i A_i \exp(-B_i), \tag{4.1}$$

where N_0 is a normalization constant, and B_i are the actions of classical solutions of the field equations which are compact and which have the given 3-metric h_{ij} and matter field configuration ϕ on the boundary. The prefactors A_i are given by the determinants of small fluctuations about the classical solutions. I shall not bother with them but only with the more important quantities B_i which appear in the exponential.

The simplest example of a minisuperspace model is that for a spatially homogeneous and isotropic universe. Such models have been considered by [10–13]. The euclidean metric can be written in the form

$$ds^2 = \sigma^2[N^2(\tau)\, d\tau^2 + a^2(\tau)\, d\Omega_3^2], \tag{4.2}$$

where $N(\tau)$ is the lapse function, $\sigma^2 = 2/(3\pi m_p^2)$ and $d\Omega_3^2$ is the metric on a 3-sphere of unit radius. The gravitational action is

$$\tilde{I} = \tfrac{1}{2} \int d\tau \left\{\frac{N}{a}\right\} \left[-\left\{\frac{a\, da}{N\, d\tau}\right\}^2 - a^2 \right]. \tag{4.3}$$

There is no solution of the classical lorentzian field equations for an empty closed homogeneous and isotropic universe. One therefore has to include some form of

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

matter. The simplest equations are obtained for a conformally invariant scalar field ϕ which is constant on the 3-sphere sections. The euclidean action of such a field is

$$\tilde{I} = \tfrac{1}{2}\int \mathrm{d}\tau \left(\frac{N}{a}\right)\left[\left(\frac{a\,\mathrm{d}\chi}{N\,\mathrm{d}\tau}\right)^2 - \chi^2\right], \tag{4.4}$$

where $\chi = \sqrt{2}/\pi a\sigma\phi$. The Wheeler–DeWitt equation is then

$$\frac{1}{2}\left(\frac{1}{a^p}\frac{\partial}{\partial a}\left[a^p\frac{\partial}{\partial a}\right] - a^2 - \frac{\partial^2}{\partial\chi^2} + \chi^2\right)\Psi[a,\chi] = 0\,. \tag{4.5}$$

The advantage of using a conformally invariant scalar field ϕ is that the Wheeler–DeWitt equations separates, i.e. there are solutions of the form

$$\Psi(a,\chi) = C(a)f(\chi)\,. \tag{4.6}$$

The function f obeys the harmonic oscillator equation

$$\frac{1}{2}\left(-\frac{\mathrm{d}^2}{\mathrm{d}\chi^2} + \chi^2\right)f = Ef. \tag{4.7}$$

It is therefore natural to take f to be the harmonic oscillator wave function with eigenvalues $E = n + \frac{1}{2}$. The disadvantage of using a conformally invariant scalar field is that the boundary condition of integrating over compact metrics uniquely picks out the ground state $n = 0$ of the harmonic oscillator [2]. The gravitational part of the wave function $C(a)$ is then almost the same as if there were no matter present. Such a wave function could not represent the universe we live in. For this reason, Hartle and I proposed in [2] that the quantum state defined by integrating over compact metrics should be regarded as the "ground state" of the universe but that the universe that we lived in did not obey these boundary conditions. Rather it was a linear superposition of other "excited state" solutions with higher values of n. This meant that one lost any ability to predict the quantum state of the universe because one had no means of determining the complex coefficients of the excited states in the superposition.

With hindsight it is clear that Hartle and I were misled by the mathematical simplicity of the conformally invariant scalar field. If one fills in a 3-sphere of radius a with a topologically trivial euclidean 4-geometry of the form (4.2), the action of the scalar field is independent of the form of $a(\tau)$. This means that there is no coupling between the scalar field and the gravitational degree of freedom which is a purely conformal one. Such a model cannot describe the observed universe in which the expansion is strongly coupled to the matter content. In order to obtain a model which corresponds to what we observe, it is necessary to introduce non-conformally invariant fields. In this section I shall describe the simplest such model, a homogeneous isotropic universe with a spatially constant massive scalar field. A future paper [8] will deal with a model in which the scalar field is replaced by an effective R^2 term generated by quantum corrections.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

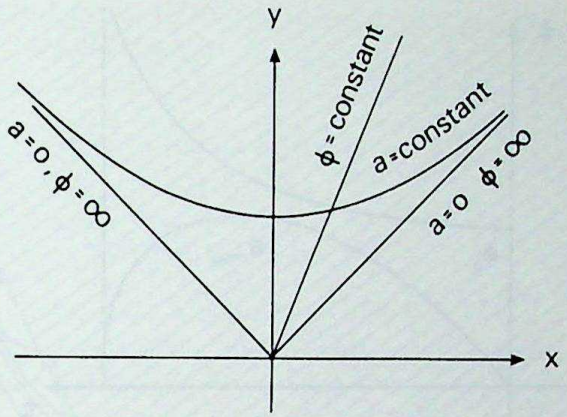


Fig. 1. The causal structure of the Wheeler–DeWitt equation can be seen in coordinates (x, y) in which the second derivatives take the form $\partial^2/\partial y^2 - \partial^2/\partial x^2$. The region $a > 0$ is mapped into the interior of the future light-cone of the origin.

The Wheeler–DeWitt equation for a homogeneous isotropic universe with a scalar field $\tilde{\phi}$ with mass $\tilde{m}$ is

$$\frac{1}{2}\left(\frac{1}{a^p}\frac{\partial}{\partial a}a^p\frac{\partial}{\partial a} - a^2 - \frac{1}{a^2}\frac{\partial^2}{\partial \phi^2} + a^4 m^2 \phi^2\right)\Psi(a,\phi) = 0\,, \tag{4.8}$$

where $\phi = \sigma\tilde{\phi}$ and $m = \sigma\tilde{m}$. The exponent p represents some, but not all of the uncertainty in the factor ordering. Its precise value will not be important for what follows. One can regard (4.8) as a wave equation in the (a, ϕ) plane. The causal structure of the Wheeler–DeWitt equation is seen more easily by introducing new coordinates

$$x = a \sinh \phi, \qquad y = a \cosh \phi\,. \tag{4.9}$$

In these coordinates, the second derivatives take the form $\partial^2/\partial y^2 - \partial^2/\partial x^2$. Thus, the characteristics are lines at 45° in the (x, y) plane. The region $a > 0$ is mapped into the interior of the future light cone of the origin (see fig. 1). The surfaces of constant ϕ are straight lines through the origin and the surfaces of constant a are spacelike hyperbolae within the light cone.

If one knew the solution on the light cone of the origin, one could solve the wave equation by standard techniques. The light cone is the surface $a = 0$, $\phi = \pm\infty$. It is therefore rather hard to apply the boundary conditions of the previous section there, but what I shall do is use the semiclassical approximation to the path integral to estimate the form of the solution on lines of constant large positive or negative ϕ.

In order to apply the semiclassical approximation, one needs to know the solutions to the classical euclidean field equations. One can express the metric in the form

$$ds^2 = \sigma^2(d\tau^2 + a^2(\tau)\, d\Omega_3^2)\,. \tag{4.10}$$

One is interested in metrics which are compact. Thus a will be zero at some value of τ which can be chosen to be zero. At $\tau = 0$, $d\phi/d\tau$ must be zero and $da/d\tau = 1$. One can then integrate the equations for $a(\tau)$ and $\phi(\tau)$ with these initial conditions

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

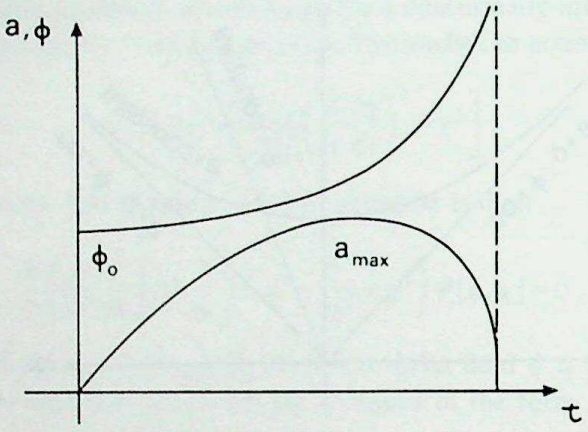


Fig. 2. Graphs of the radius a and the scalar field ϕ for a solution of the euclidean field equations.

and $\phi = \phi_0$ at $\tau = 0$. If $\phi_0 > 0$, $\phi(\tau)$ and $d\phi/d\tau$ will monotonically increase with τ and $da/d\tau$ will monotonically decrease (see fig. 2). What happens is that the positive energy density of the $m^2\phi^2(\tau)$ term in the action behaves like a positive cosmological constant and causes the 4-geometry to have positive curvature. The 3-sphere radius $a(\tau)$ rises to a maximum a_{max} which is a monotonically decreasing function of ϕ_0. The radius $a(\tau)$ then decreases and becomes zero at a singularity of the 4-geometry. If $\phi_0 \gg 1$, i.e. $\ddot{\phi} \gg m_p$, then $\phi(\tau)$ does not increase much by the time $a(\tau)$ reaches a_{max} and $a_{max} \simeq 1/m\phi_0$ (see fig. 3).

In order to estimate the wave function $\Psi(a, \phi)$, one looks for a solution $(a(\tau), \phi(\tau))$ which matches the given values of a, and ϕ at some value $\tau = \tau_0$. For sufficiently small values of a and ϕ there will be a unique trajectory $a(\tau)$, $\phi(\tau)$ which passes through the given values of a and ϕ and which does not cross any other trajectory (see fig. 3). For such a, ϕ the path integral for the wave function $\Psi(a, \phi)$ will be dominated by the contribution from the euclidean solution represented by the trajectory $a(\tau)$, $\phi(\tau)$, i.e.

$$\Psi(a, \phi) \approx N_0 \exp(-\tilde{I}[a(\tau), \phi(\tau)]) . \tag{4.11}$$

If two neighbouring trajectories $a(\tau)$, $\phi(\tau)$ in the a, ϕ plane intersect at a point a_1, ϕ_1, the path integral for the wave function $\Psi(a_1, \phi_1)$ will have a zero mode at the euclidean solution $a(\tau)$, $\phi(\tau)$. If a_2, ϕ_2 is a point on the same trajectory at a larger value of τ, the path integral for $\Psi(a_2, \phi_2)$ will have a negative mode at the euclidean solution $a(\tau)$, $\phi(\tau)$. This means that the dominant contribution to the path integral for Ψ will not come from that euclidean solution. Another way of obtaining this result is to notice that the neighbouring trajectories intersect at about the maximum value of $a(\tau)$. Beyond that value of τ, the radius a will be decreasing and the trace of the second fundamental form K will be negative. It is shown in the appendix that a solution with negative K cannot give the dominant contribution to the wave function Ψ.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

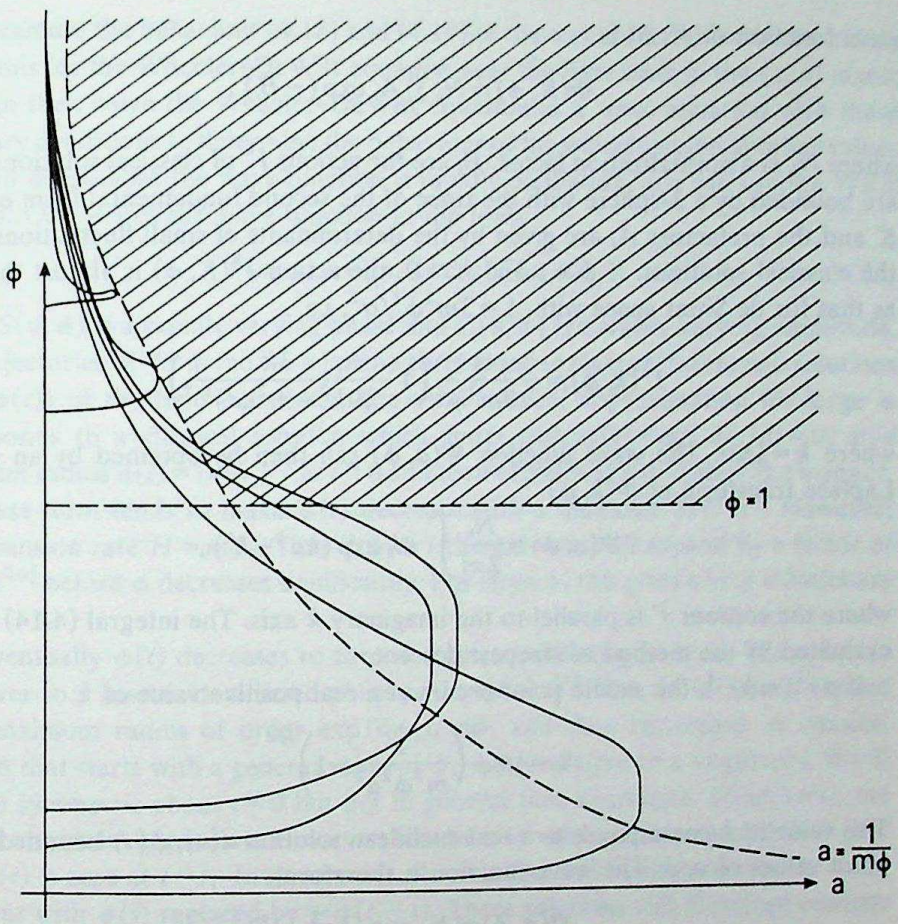


Fig. 3. Solutions of the euclidean field equations in the (a, ϕ) plane. There are no solutions which reach the region $\phi > 1$, $a < (m\phi)^{-1}$ without crossing another solution.

There is a region of the (a, ϕ) plane with $\phi > 1$, $a > (m\phi)^{-1}$ through which there are no euclidean trajectories which have not intersected another trajectory (see fig. 3).* The path integral (2.4) over compact euclidean metrics for the wave function $\Psi(a, \phi)$ for a point (a, ϕ) in that region will not receive its dominant contribution from any real euclidean metric. Instead, one can distort the contour of integration in the path integral so that the dominant contribution comes from a complex solution of the field equations. The action of such a solution will be complex and so the wave function will be oscillatory though it will be real because there will be an equal contribution from the complex conjugate solution.

One can estimate the form of the wave function in the region $\phi > 1$, $a > (m\phi)^{-1}$ by using the K representation $\Phi(K, \phi)$. The semiclassical approximation to the

* Note added in proof: D.N. Page has pointed out that there are trajectories through this region which have not intersected another trajectory. However, they correspond to solutions with large positive actions. They therefore do not give much contribution to the wave function.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

wave function $\Phi(K, \phi)$ is

$$\Phi(K, \phi) \approx N_k \sum_i A_i \exp(-B_i), \tag{4.12}$$

where N_k is a normalization factor. B_i are the actions I^k of classical solutions which are bounded by a 3-sphere with the trace of the second fundamental form equal to K and the prefactors A_i are given by the determinants of small fluctuations about the classical solutions. If $\phi \gg 1$ and $K > 0$, the action $I^k(K, \phi)$ is almost the same as that for de Sitter space with $\Lambda = 3m^2\phi^2/\sigma^2$, i.e.

$$I^k(k, \phi) \approx -\frac{1}{3m^2\phi^2}\left(1 - \frac{3k}{(9k^2 + m^2\phi^2)^{1/2}}\right), \tag{4.13}$$

where $k = \frac{1}{9}K\sigma$. The wave function $\Psi(a, \phi)$ can then be obtained by an inverse Laplace transform of $\Phi(k, \phi)$

$$\Psi(a, \phi) \approx \frac{N_k}{2\pi i}\int_\Gamma \mathrm{d}k \exp(ka^3 - I^k), \tag{4.14}$$

where the contour Γ is parallel to the imaginary k axis. The integral (4.14) can be evaluated by the method of steepest descent.

If $a < (m\phi)^{-1}$, the saddle point occurs at a real positive value of k

$$k = \tfrac{1}{3}m\phi\left(\frac{1}{m^2\phi^2a^2} - 1\right)^{1/2}. \tag{4.15}$$

This value of k corresponds to a real euclidean solution $a(\tau)$, $\phi(\tau)$ bounded by the given values of a, ϕ. The wave function is therefore

$$\begin{aligned}\Psi(a, \phi) &\approx N_k \exp(ka^3 - I^k) \\ &\approx N_k \exp(-\tilde{I}(a, \phi)),\end{aligned} \tag{4.16}$$

where $\tilde{I}(a, \phi)$ is the action of the solution bounded by the given values of a and ϕ:

$$\tilde{I}(a, \phi) = -\frac{1}{3m^2\phi^2}(1 - (1 - m^2\phi^2a^2)^{3/2}). \tag{4.17}$$

If the radius a is greater than $(m\phi)^{-1}$ there will not be a stationary phase point in (4.14) for real values of k. This corresponds to the fact that there is no real euclidean solution. However, if a is only slightly greater than $(m\phi)^{-1}$, one can analytically extend the approximate expression (4.13) for I^k and find stationary phase points at imaginary values of k:

$$k = \pm\tfrac{1}{3}im\phi\left(1 - \frac{1}{m^2\phi^2a^2}\right)^{1/2}. \tag{4.18}$$

It is possible to distort the contour Γ to pass through both of these stationary phase points. The wave function therefore has the form:

$$\Psi(a, \phi) \approx N_k \exp\left(\frac{1}{3m^2\phi^2}\right)\cos((m^2\phi^2a^2 - 1)^{3/2}/(3m^2\phi^2) - \tfrac{1}{4}\pi). \tag{4.19}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

One can use the behaviour (4.16) and (4.19) of the wave function as the boundary conditions for the Wheeler–DeWitt equation near the light cone in the (x, y) plane. One can then solve the Wheeler–DeWitt equation as a wave equation with these boundary conditions to determine the behaviour of the wave function at other values of a and ϕ. Because the wave function is rapidly oscillating, one can use the WKB method. One writes the wave function in the form

$$\Psi(a, \phi) = C(a, \phi) \cos[S(a, \phi)], \tag{4.20}$$

where $S(a, \phi)$ is a rapidly varying phase and $C(a, \phi)$ is a slowly varying amplitude. The trajectories of ∇S in the (a, ϕ) plane correspond to classical lorentzian solutions $(a(t), \phi(t))$ of the field equations. The semi-classical approximation for large ϕ corresponds to a classical solution which starts with $\phi(t) = \phi_1$, $\mathrm{d}\phi/\mathrm{d}t = 0$, at a minimum radius $a(t) = (m\phi_1)^{-1}$ at $t = 0$ and then expands in a de Sitter-like manner. The mass term tends to make $\phi(t)$ decrease with a timescale $3H/m^2$. However, the expansion rate $H = m\phi_1$. Thus if $\phi_1 > 1$ the universe will expand by a factor of order $\mathrm{e}^{3\phi_1^2}$ before ϕ decreases significantly. For large ϕ_1 this gives a long inflationary period.

Eventually $\phi(t)$ decreases to zero and starts oscillating in time. The universe goes over to a matter-dominated phase with $a(t) \propto t^{2/3}$. The universe will expand to a maximum radius of order $\exp(6\phi_1^2)/m\phi_1$ and then recollapse. A classical solution that starts with a general value of ϕ_1 will recollapse to a singularity. It will be time symmetric about $t = 0$ but not in general time symmetric about $t = t_1$, the time of maximum expansion. However, for discrete values of the scalar field ϕ_1 the field $\phi(t)$ is zero at $t = t_1$. In these cases, the solution will be time symmetric about $t = t_1$ but with $\phi(t)$ replaced by $-\phi(t_1 - t)$. These solutions will therefore oscillate indefinitely. There are also periodic solutions which are time symmetric about either the minimum or the maximum radius, but not both and there are oscillating solutions which are not time symmetric at all and which are aperiodic. The orbits of ∇S in the (a, ϕ) plane correspond to these non-singular oscillating solutions. One thus obtains a wave function $\Psi(a, \phi)$ which can be interpreted as a superposition of wave functions corresponding to oscillating universes in the classical limit. Note that the wave function does not represent a single classical universe but a whole ensemble of classical universes which expand out to different maximum radii.

The wave function $\Psi(a, \phi)$ will grow exponentially in a region near the y-axis in which $a^2 m^2 \phi^2 < 1$. This part of the wave function can be interpreted as corresponding to a flat euclidean 4-geometry. If one wishes the wave function to be normalizable, one could introduce a tiny Λ term. This would give a λa^4 term in the Wheeler–DeWitt equation (4.8) where $\lambda = \frac{1}{3}\sigma^2\Lambda$. It would cut off the exponential growth of Ψ at very large values of a. Superimposed upon this exponentially growing wave function will be an oscillating component which corresponds to the periodic solutions described above. The oscillating component in the wave function should be interpreted as corresponding to a lorentzian geometry and the exponentially growing component in the wave function should be interpreted as

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

corresponding to a euclidean geometry. We live in a lorentzian geometry and therefore we are interested really only in the oscillatory part of the wave function.

5. Beyond minisuperspace

In the previous section it was shown that the quantum state defined by the path integral over all compact euclidean metrics was a reasonable representation of the observed universe in the case of a simple minisuperspace model in which nearly all the degrees of freedom of the gravitational and matter fields were frozen out. The question naturally arises as to whether the quantum state would still be a good description of the universe if one included all those extra degrees of freedom or whether it was a misleading result produced by the truncation.

The important feature of the wave function in the model of the previous section was that it was an oscillatory rather than an exponential function of the coordinates (a, ϕ) in a certain region of minisuperspace. This allowed it to be interpreted as corresponding to lorentzian 4-geometries in the classical limit. The reasons that the wave function oscillated were:

(1) the Wheeler–DeWitt equation was a hyperbolic wave equation on minisuperspace;

(2) there was a region of minisuperspace which could not be reached by real euclidean solutions of the field equations without negative modes. This meant that the dominant contribution to the path integral for the wave function came from complex solutions of the field equations.

The "metric" $-G_{ijkl}$ in the full Wheeler–DeWitt equation that multiplies the second functional derivatives of the wave function $\partial^2\Psi/\partial h_{ij}\partial h_{kl}$ has a signature $(+------)$ at each point of the 3-manifold S. It might therefore appear that the gravitational part of the Wheeler–DeWitt equation was ultra-hyperbolic, having a signature with one plus and five minuses at each point of S. However, all but one of these infinite number of pluses can be regarded as arising from the gauge invariance of the theory. Gauge invariance manifests itself in two ways: there is the obvious freedom to make coordinate transformations in the 3-surface S. This is reflected in the equation (2.13) which states that the wave function is unchanged by coordinate transformations in S. The other gauge invariance, which is less apparent, is the freedom to choose S at different positions in the 4-geometry $g_{\mu\nu}$. Given a 4-geometry, one could, locally at least, restrict the positions of S in it to a one parameter family of non-intersecting 3-surfaces by imposing one gauge condition on the 3-metric h_{ij} such as ${}^3R = \text{constant}$. With this restriction on h_{ij}, the gravitational part of the Wheeler–DeWitt equation is hyperbolic: the one plus sign corresponds to the overall size of the 3-surface S. The matter field variables in the Wheeler–DeWitt equation all have negative signature provided that they obey reasonable positive energy conditions. Thus the full Wheeler–DeWitt equation can be regarded as a hyperbolic wave equation, in a certain sense.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

The other important property of the model of the previous section, that there should be a region of superspace or minisuperspace which cannot be reached by real euclidean solutions, seems to depend on the presence of fields with non-zero rest-mass such as a massive scalar field. Investigations of minisuperspace models containing a conformally invariant scalar field [2] or empty homogeneous anisotropic models [14] indicate that there is always a real euclidean solution of the field equations which gives the dominant contribution to the path integral for the wave function. The wave function is therefore an exponential rather than an oscillatory function of the superspace variables and so corresponds to a euclidean rather than a lorentzian 4-geometry in the classical limit.

In the massive scalar field minisuperspace model the potential $-a^2+a^4m^2\phi^2$, the non-derivative term in the Wheeler–DeWitt equation (4.8), changes sign from negative to positive on a surface in the (a, ϕ) plane which is "spacelike" for large $|\phi|$, i.e. which has a negative signature with respect to the second derivative terms in the Wheeler–DeWitt equation. Points below this surface can be reached by euclidean solutions, so the wave function is exponential in this region. The wave function on the surface $-a^2+a^4m^2\phi^2=0$ is of the form $\Psi \approx N_0 \exp(1/3m^2\phi^2)$. A solution of the wave equation with this initial data must oscillate in a region of positive potential. This is why one obtains an oscillatory wave function which can be interpreted in terms of lorentzian 4-geometries in the classical limit. This feature of the massive scalar field model does not depend on the truncation to a finite number of degrees of freedom by the imposition of symmetries: it would also be present in a model in which the 3-metric h_{ij} and the scalar field ϕ were fully general and did not have any symmetries. Thus one would expect that the solution of the Wheeler–DeWitt equation on the full superspace of the massive scalar field model would be oscillatory if it obeyed the boundary condition that it was given by a path integral over compact metrics.

Under a scale transformation $g_{\mu\nu} \to \Omega^2 g_{\mu\nu}$, $\Omega =$ constant, the gravitational action transforms as $\tilde{I} \to \Omega^2 \tilde{I}$. This means that the potential terms in the gravitational part of the Wheeler–DeWitt equation cannot change sign under a "timelike" motion in superspace that corresponds simply to a scale transformation of the 3-metric h_{ij}. In other words, the gravitational part of the potential can change sign only on a "timelike" surface in superspace. The action of a conformal invariant field, such as the electromagnetic field or a conformally invariant scalar field, is unchanged under a scale transformation. Thus the potential terms in the Wheeler–DeWitt equation arising from such fields can also change sign only on "timelike" surfaces in superspace. In fact, under reasonable positive energy assumptions, the potential terms produced by such fields will always be positive. Thus the potential cannot change from negative to positive in a model that contains only zero, rest-mass fields, either conformally invariant fields or the gravitational field. One would therefore not expect minisuperspace models that contained only zero rest-mass fields to produce oscillatory wave functions.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

On the other hand, if one includes all the infinite number of gravitational and matter degrees of freedom, one will generate R^2 and C^2 terms in the effective action. These terms will be finite in the case of a supergravity theory in which all the divergences cancel or in the case that gravity provides its own cutoff (possibilities (i) and (ii) of sect. 1) and will be infinite but renormalized in the case of a higher derivative theory (possibility (iii)). Such terms in the effective action behave like massive scalar or spin-2 particles. One might therefore hope that they would give rise to oscillatory wave functions. This has been verified, at least in the case of a minisuperspace model with an R^2 term. This will be described in a forthcoming paper [8]. It may well be therefore that the observed universe owes its existence to quantum gravitational effects.

I am grateful for discussions with J.B. Hartle, A.D. Linde, J. Luttrell and I.G. Moss.

Appendix

THE DOMINANT CONTRIBUTION TO THE PATH INTEGRAL

The wave function Ψ is given by the inverse Laplace transform

$$\Psi[h_{ij}, \phi] = \int_\Gamma \mathrm{d}\left(\frac{m_\mathrm{p}^2}{24\pi i}K\right) \exp\left(\frac{m_\mathrm{p}^2}{12\pi}\int \mathrm{d}^3x\, h^{1/2}K\right) \Phi[\tilde{h}_{ij}, K, \phi], \qquad (A.1)$$

where the wave function Φ in the K representation is

$$\Phi[\tilde{h}_{ij}, K] = \int \mathrm{d}[g_{\mu\nu}] \exp\left(-I^k[g_{\mu\nu}, \phi]\right). \qquad (A.2)$$

One expects the dominant contribution to Φ to come from metrics near a solution $\tilde{g}_{\mu\nu}$ of the field equations with the given values of $\tilde{h}_{ij}$ and K on S. The dominant contribution to Ψ will then come from the saddle point in the contour integral (A.1). At the saddle point the exponent of e in (A.1)

$$\frac{m_\mathrm{p}^2}{12\pi}\int_\mathrm{S} \mathrm{d}^3x\, h^{1/2}K - I^k[\tilde{g}_{\mu\nu}, \phi] \qquad (A.3)$$

will be an extremum under small variations of K:

$$\frac{m_\mathrm{p}^2}{12\pi} h^{1/2} = \frac{\delta I^k}{\delta K}. \qquad (A.4)$$

Eq. (A.4) implies that $\tilde{g}_{\mu\nu}$ is a solution of the field equations with the boundary conditions h_{ij}, ϕ on S. However, if $\tilde{g}_{\mu\nu}$ is a real euclidean metric, then in order for it to be a saddle point in (A.1), the exponent (A.3) must be, not only an extremum, but a minimum on the real K-axis at each point of S, i.e.

$$\frac{\delta h^{1/2}}{\delta K} < 0, \qquad (A.5)$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where $h^{1/2}$ is regarded as a function of the boundary data $(\tilde{h}_{ij}, K, \phi)$ on S that determine the solution $\tilde{g}_{\mu\nu}$.

Consider the case $\Lambda = 0$ and $\phi = 0$ on S. Under a scale transformation

$$\delta K = -\varepsilon K, \qquad \delta h^{1/2} = 3\varepsilon h^{1/2}. \tag{A.6}$$

If the 3-metric $\tilde{h}_{ij}$ is conformally flat and the 4-metric $g_{\mu\nu}$ is asymptotically euclidean, the solution $\tilde{g}_{\mu\nu}$ will be flat euclidean space outside a round 3-sphere S. For this solution K will be constant and negative. Thus under a scale transformation

$$\frac{\delta h^{1/2}}{\delta K} > 0. \tag{A.7}$$

This shows that the metric $\tilde{g}_{\mu\nu}$ cannot be a saddle point in the contour integral (A.1) and so cannot give the dominant contribution to the wave function Ψ. One can now consider a continuous variation of the values of Λ and the matter field configuration ϕ on S from zero. Under such a variation the left-hand side of (A.7) will change but it cannot become negative everywhere on S unless there is some Λ and ϕ for which the left-hand side of (A.7) is either infinite or zero everywhere on S. In the first case, the solution $\tilde{g}_{\mu\nu}$ will have a zero mode for fixed $\tilde{h}_{ij}$, K and ϕ on S, and in the second case it will have a zero mode for fixed h_{ij} and ϕ. If one continues the variation until the left side of (A.7) is negative, the solution $\tilde{g}_{\mu\nu}$ will have a negative mode for the given boundary conditions. In the first case, this indicates that $\tilde{g}_{\mu\nu}$ does not provide the dominant contribution to the wave function Φ and hence does not provide the dominant contribution to Ψ. In the second case, the direct representation of Ψ in terms of a path integral shows that $\tilde{g}_{\mu\nu}$ does not provide the dominant contribution to Ψ.

References

[1] S.W. Hawking, in Astrophysical cosmology, Pontificiae Academiae Scientarium Scripta Varia, 48 (1982) 563
[2] J.B. Hartle and S.W. Hawking, Phys. Rev. D28 (1983) 2960
[3] A.H. Guth, Phys. Rev. D23 (1981) 347
[4] A.D. Linde, Phys. Lett. 108B (1982) 389
[5] S.W. Hawking and I.G. Moss, Phys. Lett. B110 (1982) 35
[6] A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. 48 (1982) 120
[7] A.H. Guth and S.Y. Pi, Phys. Rev. Lett. 49 (1982) 1110
[8] S.W. Hawking and J. Luttrell, paper in preparation
[9] S.W. Hawking, Quantum cosmology, Les Houches lectures (1983), to be published
[10] B.S. DeWitt, Phys. Rev. 160 (1967) 113
[11] C.J. Isham and J.E. Nelson, Phys. Rev. D10 (1974) 3226
[12] W.E. Blyth and C.J. Isham, Phys. Rev. D11 (1975) 768
[13] I.G. Moss and W.A. Wright, The wave function of the inflationary universe, Newcastle Univ. preprint (1983)
[14] J. Luttrell, private communication

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Nuclear Physics B247 (1984) 250–260
© North-Holland Publishing Company

HIGHER DERIVATIVES IN QUANTUM COSMOLOGY
(I). The isotropic case

S.W. HAWKING and J.C. LUTTRELL

University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, UK

Received 2 July 1984

This paper is the first in a series that investigates the effects on cosmology of curvature-squared terms which are bound to appear in the effective action whether or not gravity is perturbatively finite, is cut off by nonperturbative effects, or is made renormalisable by the addition of curvature squared terms to the bare action. We examine how these terms affect the recent proposal that the quantum state of the universe is defined by a path integral over compact metrics. In this paper we consider a simple minisuperspace model of an isotropic universe. In such a model the C-squared term in the action plays no role while the R-squared term behaves like a massive scalar field. The wave function of the universe can be interpreted as corresponding in the classical limit to a family of solutions that start out with a long period of exponential expansion and then go over to a matter dominated era.

1. Introduction

The quantum state of the universe may be described by a wave function Ψ which is a functional of the metric h_{ij} on a three-dimensional surface S and of the matter field configurations ϕ on S. The wave function obeys a couple of equations which are obtained from the hamiltonian and momentum constraint equations of classical general relativity by replacing momenta by functional derivatives with respect to the conjugate quantities:

$$\left\{-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}}+h^{1/2}\left[-{}^3R+2\Lambda+\frac{1}{16\pi}T_{00}\left(\frac{\delta}{\delta\phi},\phi\right)\right]\right\}\Psi=0, \tag{1.1}$$

where

$$G_{ijkl}=\tfrac{1}{2}h^{-1/2}\left(h_{ik}h_{jl}+h_{il}h_{jk}-h_{ij}h_{kl}\right), \tag{1.2}$$

$$\left(\frac{\delta\Psi}{\delta h_{ij}}\right)\Bigg|_i=\tfrac{1}{2}T^{0j}. \tag{1.3}$$

Any solution of (1.1) and (1.3) represents a possible quantum state for the universe. It would seem reasonable to suppose that the solution which corresponds

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

to the actual quantum state of the universe should be picked out in some special way. It has recently been suggested [1–4] that the most natural choice of quantum state for the universe is that given by a path integral over compact metrics

$$\Psi\left[h_{ij}, \phi_0\right] = \int \mathrm{d}\left[g_{\mu\nu}\right] \mathrm{d}\left[\phi\right] \exp\left(-\hat{I}\left[g_{\mu\nu}, \phi\right]\right), \tag{1.4}$$

where $\hat{I}$ is the euclidean action

$$\hat{I} = \frac{-1}{16\pi} \int \mathrm{d}^4x \sqrt{g}\,(R - 2\Lambda) - \frac{1}{8\pi} \int \mathrm{d}^3x \sqrt{h}\,K - \int \mathrm{d}^4x \sqrt{g}\,L_m, \tag{1.5}$$

and the integral is taken over all compact positive definite metrics and matter field configurations which have S as their boundary on which the induced metric is h_{ij} and matter field configuration is ϕ_0. It may be verified that the wave function given by (1.4) is a solution to (1.1) and (1.3).

In order to test this suggestion for the quantum state one would like to calculate the wave function, compute the expectation values of observable quantities and compare them with the results of observations. In order to calculate the wave function one has to solve eqs. (1.1) and (1.3) which are functional differential equations on an infinite-dimensional space called superspace, the space of all 3-metrics h_{ij} and matter field configurations of ϕ_0 on S. This is rather difficult but one can hope to get some idea of the quantum state by solving (1.1) and (1.3) on a finite-dimensional approximation to the full superspace. In other words one restricts the degree of freedom of the gravitational and matter fields down to a finite number and solves (1.1) and (1.3) as partial differential equations on a finite-dimensional manifold called minisuperspace. This has been done for a homogeneous, isotropic universe in refs. [3] and [4], and for a homogeneous anisotropic universe in ref. [5]. In these examples it was found necessary to include a field of nonzero rest mass in order to get a wave function which oscillated as a function of the radius of the universe and so corresponded to a lorentzian rather than a euclidean 4-geometry in the classical limit.

The examples described above used the simplest case of a field of nonzero rest mass, a massive, minimally coupled scalar field. This led to wave functions which could be interpreted via the WKB approximation as corresponding to classical lorentzian solutions of the field equations which had a period of exponential expansion and which approached isotropy at the present epoch. They were thus in agreement with observations.

The requirement to have a field of nonzero rest mass does not seem to depend on the symmetry of these models. A massive scalar field was convenient because of its mathematical simplicity but there are grounds for doubting whether there really are any scalar fields which are not bound states of fields of higher spin. One would

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

therefore prefer an explanation of the observed wave function of the universe which did not depend in an essential way on the problematical existence of a fundamental massive scalar field. Fortunately, it seems that quantum corrections to the gravitational action can provide an alternative mechanism for generating fields with nonzero rest mass.

As is well known, the usual perturbation expansion applied to a system of interacting gravitational and matter fields gives rise to divergences which are proportional to terms which are quadratic or higher in the curvature tensor. These infinities cannot be absorbed by the usual procedure of renormalisation of the terms in the action because such terms do not appear in the original action. There are three possible solutions to this problem:

(i) we may be able to find a theory, such as $N = 8$ supergravity, in which these divergences cancel to all orders.

(ii) the divergences may simply be an artifact of perturbation theory which is known to break down at high energies; one cannot represent a black hole as a perturbation of flat space.

(iii) if neither of the above works, one will have to add terms proportional to the curvature squared to the action. This would lead to a renormalisable theory [6].

Whichever of these is the correct solution, the effective action including quantum correction will contain terms proportional to the curvature squared. These terms will be finite in the case of possibilities (i) and (ii) and will be infinite but renormalised in the case of possibility (iii). Perturbations around flat space indicate that such quadratic curvature terms give rise to massive scalar and spin-two particles. One might therefore hope that they would lead to a wave function which corresponded to the observed universe. This is what we shall investigate in this and subsequent papers. Of course the effective action will also contain terms which are cubic or of higher order in the curvature as well as nonlocal terms. However, one would hope that these other terms could also be interpreted in terms of additional massive fields. In this and the other papers in this series we shall consider only the quadratic terms.

There are four independent scalar polynomials which are quadratic in the curvature. There are, however, two topological invariants, the Hirzebrucht signature and the Euler number, which can be expressed as integrals of linear combinations of these polynomials. Thus if one considers the path integral expression (1.4) restricted to four-geometries of a given topology, there are only two independent coupling constants for the quadratic terms in the effective euclidean action. We shall take them to be of the form

$$\hat{I} = \frac{-1}{16\pi}\int \mathrm{d}^4x\sqrt{g}\left(R - \alpha C^{\mu\nu\rho\sigma}C_{\mu\nu\rho\sigma} + \beta R^2\right)$$

$$-\frac{1}{8\pi}\int \mathrm{d}^3x\sqrt{h}\left(K(1+2\beta R) - 4\alpha K_{ij}C^{\mu i\nu j}n_\mu n_\nu\right), \qquad (1.6)$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where $C_{\alpha\beta\gamma\delta}$ is the Weyl tensor and n^{α} is the unit normal to the boundary. Perturbations about flat space indicate that the R^2 term behaves like a scalar particle with $\text{mass}^2 = (6\beta)^{-1}$. The C^2 term behaves like a spin-two particle with $\text{mass}^2 = 2/\alpha$. However, if α is positive, as is required for a positive mass squared, the kinetic energy of the massive spin-two particle appears in the action with the wrong sign. This is normally held to lead to problems with negative energy states, instabilities and loss of unitarity in particle scattering calculations in asymptotically flat space. However it is an observational fact that our universe is not asymptotically flat: there is enough matter on our past light cone to cause it to refocus [7]. The total energy of the universe is exactly zero, the positive energy of the gravitons and matter particles being exactly compensated by the negative gravitational potential energy: that is why the universe is expanding. Finally, it is not clear that unitarity has any well-defined meaning except in scattering calculations in asymptotically flat spaces. We therefore see no reason to be frightened of C^2 terms in the action. Indeed such terms will be there whichever of the three possibilities listed above is correct. One therefore has to learn to live with them.

In order to derive the equations analogous to (1.1) and (1.3) for the wave function, one has to find a canonical form for the higher-derivative theory. This has been studied by Boulware [8]. One treats as coordinates not only the six components of the three-metric h_{ij} on S but also the five components E_{ij} of the electric part of the Weyl tensor and the four-dimensional scalar curvature R. The second fundamental form K_{ij} of S is then the momentum conjugate to E_{ij} and R while the momentum conjugate to h_{ij} involves the time derivatives of E_{ij} and R. One obtains four constraint equations involving these coordinates and their conjugate momenta. The quantum theory is obtained by replacing the momenta by derivatives with respect to the conjugate coordinates in the normal way. One then looks for a wave function of h_{ij}, E_{ij} and R which satisfies these equations and obeys the boundary condition that it is given by a path integral over compact metrics.

In this paper we shall consider only the simplest minisuperspace example, that of spatially homogeneous and isotropic universe. The four-geometries with this symmetry are all conformally flat. This means that the αC^2 term in the action (1.6) is zero and can be neglected for the particular solution that we are considering. The wave function Ψ will be a function only of the radius a of the three-sphere and the four-curvature R. We shall show that it behaves very similarly to the wave function of the model with the ordinary Einstein action and a massive scalar field. It also corresponds to a family of lorentzian solutions of the classical field equations which have a long period of exponential expansion and which could represent the observed universe.

The C^2 term in the action would play a role in models that were not restricted to be exactly homogeneous and isotropic. This will be investigated in subsequent papers. However the results obtained here indicate that there is at least a neighbourhood of homogeneity and isotropy in which higher-derivative actions can lead to a

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

reasonable wave function for the universe. Whether this neighbourhood is large enough to accommodate and explain the inhomogeneities that we observe remains to be seen.

2. Canonical formalism

We shall consider four-geometries of the form

$$ds^2 = \frac{4}{3\pi}\left(d\tau^2 + a^2 d\Omega_3^2\right), \tag{2.1}$$

where $d\Omega_3^2$ is the metric on a unit three-sphere. It is helpful to introduce a new time coordinate

$$\eta = \int \frac{d\tau}{a}, \tag{2.2}$$

so that the metric takes the form

$$ds^2 = \frac{4a^2}{3\pi}\left(d\eta^2 + d\Omega_3^2\right). \tag{2.3}$$

The action is then

$$\hat{I} = -\int d\eta \left(\frac{1}{a^2} - \frac{a''}{a^3} + \tfrac{9}{2}\pi\beta\left(\frac{1}{a^2} - \frac{a''}{a^3}\right)^2\right) a^4, \tag{2.4}$$

where $a' = da/d\eta$. We can express the action as a function of a and R or, equivalently, as a function of a and $Q = a(1 + 2\beta R)$ and their first derivatives with respect to η:

$$\hat{I} = -\int d\eta\, L = \int d\eta \left[-Q'a' + \frac{Q^2a^2}{2\tilde{\beta}} - \frac{Qa^3}{\tilde{\beta}} - Qa + \frac{a^4}{2\tilde{\beta}}\right], \tag{2.5}$$

where $\tilde{\beta} = 9\pi\beta$. One can then treat a and Q as independent coordinates and define their canonical momenta in the usual way:

$$P_a = \frac{\partial L}{\partial a'} = Q',$$

$$P_Q = \frac{\partial L}{\partial Q'} = a'. \tag{2.6}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

One can then define the hamiltonian by the normal Legendre transformation:

$$H = P_a a' + P_Q Q' - L$$

$$= +P_a P_Q + \frac{Q^2 a^2}{2\tilde{\beta}} - \frac{Qa^3}{\tilde{\beta}} - Qa + \frac{a^4}{2\tilde{\beta}}. \tag{2.7}$$

One can also express the hamiltonian in terms of coordinates x and y:

$$x = Q - a, \qquad y = Q + a, \tag{2.8}$$

$$H = -\tfrac{1}{4}P_x^2 + \tfrac{1}{4}P_y^2 - \tfrac{1}{4}(y^2 - x^2) + \frac{x^2}{8\tilde{\beta}}(x-y)^2. \tag{2.9}$$

The classical equations of motion follow in the normal way:

$$x' = \frac{\partial H}{\partial P_x} = -\tfrac{1}{2}P_x, \tag{2.10a}$$

$$y' = \frac{\partial H}{\partial P_y} = \tfrac{1}{2}P_y, \tag{2.10b}$$

$$P_x' = -\frac{\partial H}{\partial x} = -\tfrac{1}{2}x\left[1 + \frac{1}{2\tilde{\beta}}(2x^2 - 3xy + y^2)\right], \tag{2.10c}$$

$$P_y' = -\frac{\partial H}{\partial y} = \tfrac{1}{2}y - \frac{x^2}{4\tilde{\beta}}(y-x). \tag{2.10d}$$

The quantum theory is obtained by replacing P_x and P_y by $-\partial/\partial x$ and $-\partial/\partial y$ respectively. There is no factor of i because we are dealing with euclidean momenta. There are ambiguities in the factor ordering but these will not be important in our case. We shall therefore take the simplest ordering. The wave function thus satisfies the equation

$$\left[\frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial x^2} + 4V(x,y)\right]\Psi = 0, \tag{2.11}$$

where

$$V = \frac{x^2}{8\tilde{\beta}}(y-x)^2 - \tfrac{1}{4}(y^2 - x^2). \tag{2.12}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

3. The field equations

The classical field equations that are derived from the action

$$\hat{I} = -\frac{1}{16\pi}\int \mathrm{d}^4x\sqrt{g}\,(R+\beta R^2) + \text{surface terms}, \tag{3.1}$$

are [9]

$$R_{\mu\nu}(1+2\beta R) = \tfrac{1}{2}g_{\mu\nu}\beta R^2 + 2\beta R_{;\mu\nu} + \tfrac{1}{2}g_{\mu\nu}(R - 4\beta\Box R). \tag{3.2}$$

The trace of this equation is

$$\Box R - \frac{1}{6\beta}R = 0. \tag{3.3}$$

This shows that R behaves like a scalar field with $m^2 = 1/6\beta$. Whitt [10] has shown that one can express the full fourth-order field equations (3.2) in a manner which resembles ordinary Einstein gravity coupled to a massive scalar field. Define a new metric by

$$\tilde{g}_{\mu\nu} = (1+2\beta R)\,g_{\mu\nu}. \tag{3.4}$$

Then

$$\tilde{R}_{\mu\nu} = R_{\mu\nu} - \frac{2\beta R_{;\mu\nu} + g_{\mu\nu}\beta\Box R}{1+2\beta R} + \frac{6\beta^2 R_{;\mu}R_{;\nu}}{(1+2\beta R)^2}, \tag{3.5}$$

where $\phi = (3/4\pi)^{1/2}\beta R$. Thus

$$\tilde{R}_{\mu\nu} - \tfrac{1}{2}\tilde{g}_{\mu\nu}\tilde{R} = 8\pi\left(1+4\left(\tfrac{1}{3}\pi\right)^{1/2}\phi\right)^{-2}\left[\phi_{;\mu}\phi_{;\nu} - \tfrac{1}{2}\tilde{g}_{\mu\nu}\left(\phi_{;\lambda}\phi^{;\lambda} + \frac{1}{6\beta}\phi^2\right)\right]. \tag{3.6}$$

Eq. (3.6) is the Einstein equation for the metric $\tilde{g}_{\mu\nu}$ coupled to the scalar field ϕ which has the standard energy momentum tensor multiplied by $(1+4(\frac{1}{3}\pi)^{1/2}\phi)^{-2}$. The scalar field ϕ obeys the equation which can be derived by taking the covariant divergence of (3.6):

$$\left(1+4\left(\tfrac{1}{3}\pi\right)^{1/2}\phi\right)\Box\phi - 4\left(\tfrac{1}{3}\pi\right)^{1/2}\phi_{;\nu}\phi_{;\mu}\tilde{g}^{\mu\nu} - \frac{1}{6\beta}\phi = 0. \tag{3.7}$$

When ϕ is small eqs. (3.6) and (3.7) are almost the usual equations for gravity coupled to a scalar field with $m^2 = 1/6\beta$. When $\phi = -\frac{1}{4}(3/\pi)^{1/2}$, the conformal transformation relating $\tilde{g}_{\mu\nu}$ and $g_{\mu\nu}$ is singular. This value of ϕ corresponds to $Q = 0$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

in the previous section. However the treatment given there shows that the field equations for $g_{\mu\nu}$ remain well behaved there.

The field equations (3.6) and (3.7) can be derived by varying $\tilde{g}_{\mu\nu}$ and ϕ as independent variables in the action

$$\hat{I} = -\frac{1}{16\pi}\int \mathrm{d}^4x\sqrt{\tilde{g}}\,\tilde{R} + \int \mathrm{d}^4x\sqrt{\tilde{g}}\left(\phi_{;\mu}\phi_{;\nu}\tilde{g}^{\mu\nu} + \frac{1}{6\beta}\phi^2\right)\left(1 + 4\left(\tfrac{1}{3}\pi\right)^{1/2}\phi\right)^{-2}. \tag{3.8}$$

This action is equal to the original action (3.1) apart from a total divergence. However this form is not so suitable for quantization because of the possibility that the conformal transformation may be singular.

As in the standard scalar field case there will be lorentzian solutions which start from a minimum radius with a large positive value ϕ_1 of ϕ and a small time derivative $\mathrm{d}\phi/\mathrm{d}t$. The ϕ^2 term in the action (3.5) will act like a cosmological constant $\Lambda = \frac{1}{4}\beta$. This will cause the universe to exponentially expand with $H = \dot{a}/a = (12\beta)^{-1/2}$. The scalar field will decrease linearly at the rate $\dot{\phi} = -1/12\sqrt{\pi\beta}$. The solution will inflate by a factor of order $\exp(12\sqrt{\pi}\,\beta\phi_1)$. The scalar field will then pass through zero and will oscillate with period $2\pi\sqrt{6\beta}$. The universe in the $\tilde{g}_{\mu\nu}$ metric will then expand like a matter dominated universe with a proportional to $t^{2/3}$ up to some maximum radius of the order of $\exp(4\sqrt{\pi}\,\beta\phi_1)$. In the physical $g_{\mu\nu}$ metric this matter dominated era appears as a rapidly oscillating radius superimposed on an overall expansion (fig. 1: we are grateful to Mr. Z.C. Wu for this). In a more realistic model which contained nonconformally invariant matter fields, the oscillations in

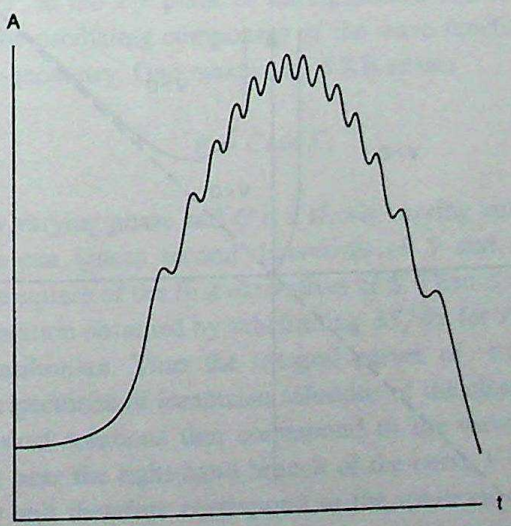


Fig. 1. The radius a of a homogeneous isotropic solution of the field equations against time t.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

the radius would cause particle creation and lead to a radiation dominated expansion with a proportional to $t^{1/2}$.

4. The wave function

Eq. (2.11) is a two-dimensional wave equation with potential V. We need to solve it in the region $a > 0$, i.e. $y > x$. We want to find the solution that has the approximate form

$$\Psi = A\,\mathrm{e}^{-B}, \tag{4.1}$$

given by the semiclassical approximation to the path integral (1.4) over compact metrics. Here B is the action of a compact euclidean solution of the classical equations bounded a three-sphere with the given values of a and Q. The prefactor A represents the contributions of metrics near the classical solution. It will depend on the measure in the path integral which will also determine the factor ordering in eq. (2.11). However this dependence will not affect our considerations because, in general, A will be an algebraic function of a and Q and so will not affect whether Ψ is exponential or oscillatory, i.e. whether it corresponds to a euclidean or lorentzian four-geometry in the classical limit.

Compact euclidean solutions of the classical field equations start at the origin of the x-y plane (fig. 2) with $a = 0$, and Q/a finite. The solutions that intersect the right-hand branch of the curve $V = 0$ have a maximum value of a there. Points in the region below the right-hand branch of $V = 0$ but above the line $a = 0$ ($y = x$)

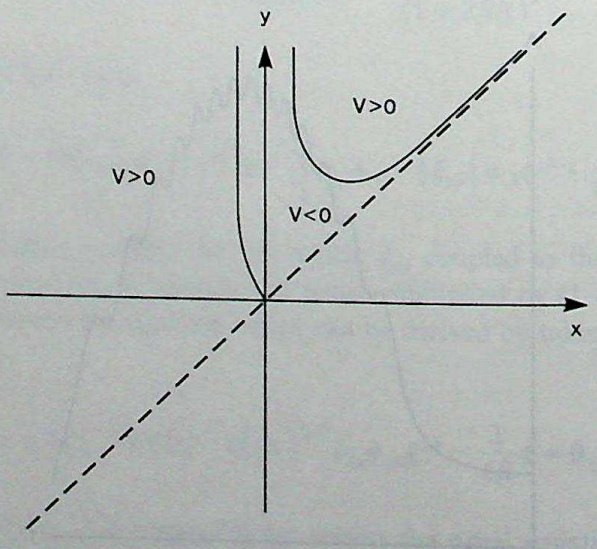


Fig. 2. Eq. (2.11) has to be solved in the region $y > x$, $a > 0$.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

can be reached by solutions of very small action. With the factor ordering choice in (2.11) the prefactor A can be taken to be 1 in this region. Thus the boundary condition for eq. (2.11) is $\Psi = 1$ on the null line $y = x$.

The line $y = x$ is not a Cauchy surface on its own: it has to be supplemented by a knowledge of Ψ and its first derivatives on a spacelike curve through the origin. We can take this curve to be the trajectory that starts at the origin with initial values $Q/a = -1$. This curve will bend downwards in the x-y plane but it will remain spacelike and in the region $Q < 0$. The action of a solution is given by (2.5). The hamiltonian constraint $H = 0$ implies

$$Q'a' = -\frac{Q^2a^2}{2\tilde{\beta}} + \frac{Qa^3}{\tilde{\beta}} + Qa - \frac{a^4}{2\tilde{\beta}}. \tag{4.2}$$

Thus on a solution

$$\hat{I} = \int \mathrm{d}\eta \left(\frac{Q^2a^2}{\tilde{\beta}} - \frac{2Qa^3}{\tilde{\beta}} - 2Qa + \frac{a^4}{\tilde{\beta}} \right). \tag{4.3}$$

Thus if Q is negative, the action increases monotonically along the trajectory of a solution. By the semiclassical approximation (4.1) Ψ and its first derivative will decrease exponentially to zero.

One can integrate the wave equation (2.11) with these boundary conditions. The wave function will grow exponentially in the region where V is less than zero (fig. 2). It will start to oscillate in the region above the right-hand branch of the curve $V = 0$. This has been confirmed by numerical calculations [11]. The oscillations will spread to the whole "future" in the x-y plane of the right-hand branch of the curve $V = 0$. One can interpret the oscillating component of the wave function as corresponding to a lorentzian four-geometry. One makes the WKB ansatz

$$\Psi = C \cos S, \tag{4.4}$$

where S is a rapidly varying phase and C is a slowly varying amplitude. In the wave function (2.11) one can ignore second derivatives of S and derivatives of C in comparison with the square of the first derivatives of S. Then S satisfies the classical Hamilton-Jacobi equation obtained by substituting $\partial S/\partial x$ for P_x and $\partial S/\partial y$ for P_y in the classical hamiltonian. Thus the integral curves of ∇S in the x-y plane correspond to the trajectories of lorentzian solutions of the classical field equations. The particular classical solutions that correspond to the wave function described above will start out near the right-hand branch of the curve $V = 0$ with $\mathrm{d}a/\mathrm{d}t$ and $\mathrm{d}R/\mathrm{d}t$ small. They will therefore correspond to the solutions described in the last section which have a long period of exponential expansion and then go over to a matter dominated expansion. They will reach a maximum radius and then recol-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

lapse. Although the general classical solution will collapse to a singularity of zero radius, the particular class of solutions that correspond to the wave function will have a bounce at or near the right-hand branch of the curve $V = 0$. At $V = 0$ however the WKB approximation will break down and the concept of time will become ill-defined. It does not really have any meaning therefore to say whether or not the universe collapses to zero radius but the wave function remains regular at $a = 0$.

Future papers in this series will investigate the behaviour of models that are not constrained to be perfectly isotropic. However the results presented here show that the wave function given by a path integral over compact metrics is reasonable in some neighbourhood of perfect isotropy. It remains to be seen whether the predicted fluctuations from isotropy and homogeneity agree with observations.

References

[1] S.W. Hawking, Pontificae Academiae Scientarum., Scripta Varia 46 (1982) 563
[2] J.B. Hartle and S.W. Hawking, Phys. Rev. D28 (1983) 2960
[3] S.W. Hawking, Lectures delivered at Les Houches Summer School, 1983, Relativity, groups and topology (North-Holland, to be published)
[4] S.W. Hawking, Nucl. Phys. B239 (1984) 257
[5] S.W. Hawking and J.C. Luttrell, The isotropy of the universe, Phys. Lett. B., to be published
[6] K.S. Stelle, Phys. Rev. D16 (1977) 953
[7] S.W. Hawking and R. Penrose, Proc. Roy. Soc. A314 (1970) 529
[8] D. Boulware, Quantum theory of gravity, ed. S.M. Christensen (Adam Hilger, 1984)
[9] N.H. Barth and S.M. Christensen, Phys. Rev. D28 (1983) 1876
[10] B. Whitt, Phys. Lett. 145B (1984) 176
[11] S.W. Hawking and Z.C. Wu, Numerical calculations of minisuperspace cosmological models, Phys Lett. B, to be published

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

NUMERICAL CALCULATIONS OF MINISUPERSPACE COSMOLOGICAL MODELS

S.W. HAWKING and Z.C. WU
University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, USA

Received 17 September 1984

The wave function of the universe is calculated numerically by solving the zero energy Schrödinger equation for homogeneous isotropic cosmological models with a massive scalar field or with curvature squared corrections to the action. The boundary condition used to integrate the equation is that the wave function should be given by a path integral over compact metrics.

The quantum state of the universe can be described by a wave function Ψ which is a functional of the three-metric h_{ij} and the matter field configuration ϕ_0 on a spacelike three-surface S. The wave function obeys the Wheeler–DeWitt or zero energy Schrödinger equation:

$$\{-G_{ijkl}\,\delta^2/\delta h_{ij}\delta h_{kl} + h^{1/2}[-{}^3R + 16\pi T_{00}(\delta/\delta\phi, \phi)]\}\,\Psi[h_{ij}, \phi] = 0, \quad (1)$$

where

$$G_{ijkl} = \tfrac{1}{2}h^{-1/2}(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}), \quad (2)$$

and the momentum constraints

$$(\delta\Psi/\delta h_{ij})_{|i} = 8\pi T^{0j}\Psi, \quad (3)$$

Any solution of eqs. (1) and (3) corresponds to a possible quantum state for the universe. However, it seems reasonable to suppose that the quantum state of *our* universe is not just arbitrary but that it is characterized in some special way. It has recently been proposed [1–4] that the quantum state of the universe is determined by a path integral over compact euclidean metrics: the wave function would be

$$\Psi[h_{ij}, \phi_0] = \int \mathrm{d}[g_{\mu\nu}]\,\mathrm{d}[\phi]\,\exp(-\hat{I}[g_{\mu\nu}, \phi]), \quad (4)$$

where $\hat{I}$ is the euclidean action

$$\hat{I} = -\frac{1}{16\pi}\int \mathrm{d}^4x\,\sqrt{g}\,R - \frac{1}{8\pi}\int \mathrm{d}^3x\,\sqrt{h}\,K - \int \mathrm{d}^4x\,\sqrt{g}\,L_{\mathrm{m}},$$

the three-manifold S as their boundary on which the induced metric is h_{ij} and the matter field configuration is ϕ_0. It can be verified that the wave function defined by (4) automatically satisfies eqs. (1) and (3) [2–4].

The path integral prescription equation (4) for the wave function can be interpreted as a boundary condition on the Wheeler–DeWitt equation which can be regarded as a second order "hyperbolic" functional differential equation on an infinite dimensional space called superspace, the space of all three metrics h_{ij} and matter field configurations ϕ_0 on S. It would be very difficult to solve this equation on such an infinite dimensional space but one could hope to get some idea of the nature of the solutions by solving the Wheeler–DeWitt equation on a finite dimensional approximation to superspace. One restricts the infinite number of degrees of freedom of the gravitational and matter fields down to a finite number and solves the Wheeler–DeWitt equation as a wave equation on a finite dimensional space called minisuperspace. Models of this kind with the boundary conditions implied by (4) have been considered in refs. [2–6]. In the interesting cases, the Wheeler–DeWitt equation could not be solved ex-

0370-2693/85/$ 03.30 © Elsevier Science Publishers B.V. (North-Holland Physics Publishing Division)

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

actly by separation of variables. Various approximations, therefore, had to be used. These suggested that the wave function had a certain qualitative behaviour and that it was oscillatory in certain regions. The oscillating component of the wave function could be interpreted via the WKB approximation as corresponding to a family of lorentzian solutions to the classical field equations in the classical limit [3–6].

The aim of this paper is to check the qualitative behaviour of the wave function by numerical integration of the Wheeler–DeWitt equation and to investigate the classical solutions to which it corresponded. We shall study two two-dimensional minisuperspace models. Both correspond to a homogeneous isotropic closed universe with metric

$$ds^2 = [2a^2(t)/3\pi](-dt^2 + d\Omega_3^2), \tag{5}$$

where $d\Omega_3^2$ is the metric on a unit three-sphere. In the first case [3,4] the model contains a massive scalar field ϕ and in the second case [6] the action contains curvature squared quantum corrections that act rather like a massive scalar field.

For the massive scalar field model the Wheeler–DeWitt equation is

$$\tfrac{1}{2}[(1/a^p)(\partial/\partial a)\, a^p\, \partial/\partial a - a^2 - a^{-2}\, \partial^2/\partial\phi^2 + a^4 m^2 \phi^2]\Psi(a,\phi) = 0. \tag{6}$$

The index p represents some of the ambiguity of the operator ordering in the Wheeler–DeWitt equation. Its value does not seem to affect such qualitative features of the wave function such as whether or not it oscillates. For convenience we shall assume $p = 1$. Eq. (6) can be put in a more convenient form by introducing new coordinates

$$x = a \sinh\phi, \quad y = a \cosh\phi. \tag{7}$$

It then becomes

$$\tfrac{1}{2}[\partial^2/\partial y^2 - \partial^2/\partial x^2 + V(x,y)]\Psi = 0, \tag{8}$$

where

$$V(x,y) = x^2 - y^2 + m^2(y^2 - x^2)^2 \operatorname{arctanh}^2(x/y).$$

One has to solve this equation for $a > 0$ which corresponds to the region $y > |x|$ in the interior of the future light cone of the origin (fig. 1). The solution will be determined by a knowledge of Ψ on or near the light cone $y = |x|$. For this one can use the semi-classical approximation to the path integral (4):

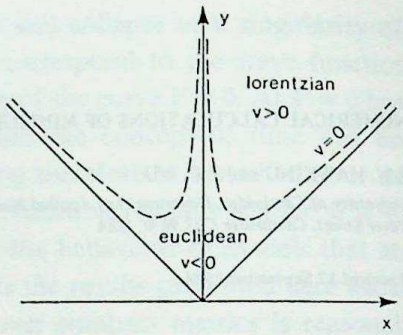


Fig. 1. The x, y plane for the massive scalar field model. The boundary condition is $\Psi = 1$ on the future light cone of the origin.

$$\Psi = A \exp(-B), \tag{9}$$

where B is the action of a compact positive definite solution of the classical field equations which is bounded by a three-sphere with the given values of a and ϕ. For a point near the light cone, the action B is very small. The behaviour of the prefactor A depends on the measure in the path integral or, equivalently, on the factor ordering in the Wheeler–DeWitt equation. If $p = 1$, it is consistent to take $A = 1$ near the light cone. Thus the boundary condition for the Wheeler–DeWitt equation implied by eq. (4) is $\Psi = 1$ on the light cone, $y = |x|$.

Eq. (8) was integrated numerically with this boundary condition using an algorithm that has been described elsewhere [7,8]. It was found that Ψ increased exponentially in the region of the (x,y) plane in which $V < 0$, thus confirming the approximations in refs. [3, 4]. In the region $V > 0$ (fig. 1) the wave function started to oscillate as predicted in refs. [3,4] and these oscillations spread to the central channel in which $V < 0$. Fig. 2 shows a computer plot of Ψ as a function of x and y for $x > 0, y > x$ in the case $m = 5$. The wave function is symmetric under $x \to -x$. In order to keep the plot from blowing up Ψ has been multiplied by

$$\exp\{-3m^{-2}\phi^{-2}[1 - (1 - m^2\phi^2 a^2)^{3/2}]\} \tag{10}$$

in the region $V < 0$, and by

$$\exp(-3m^{-2}\phi^{-2}) \tag{11}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

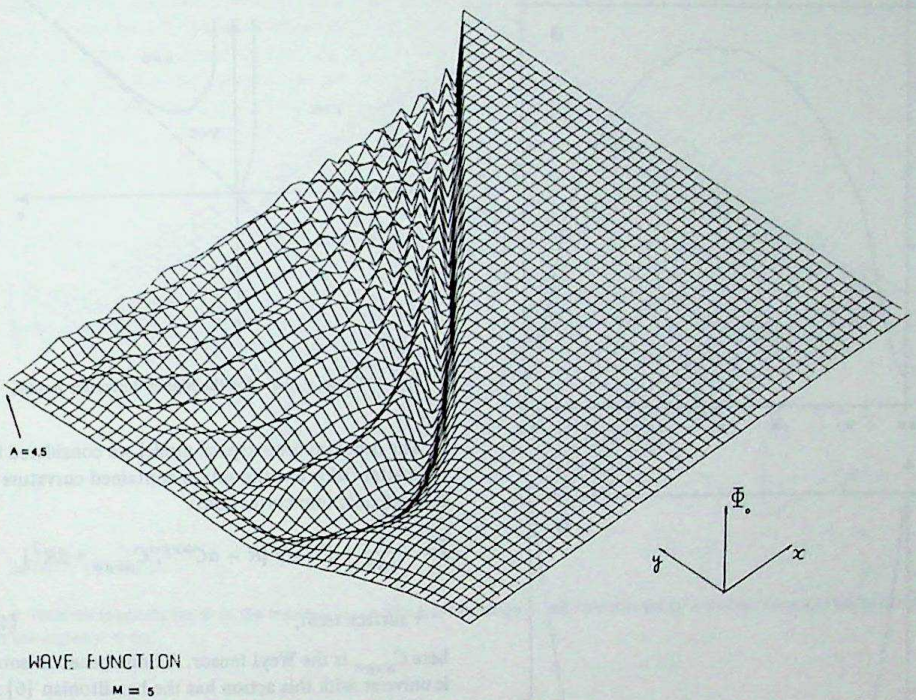


Fig. 2. Numerical results for Ψ in the region $x > 0$. To improve legibility the wave function has been divided by a function described in the text.

in the region $V > 0$. The integration had to be stopped at $y = 4.5$ because Ψ was diverging exponentially as $\exp(\frac{1}{2}a^2)$.

The oscillating component of the wave function can be interpreted by the WKB ansatz:

$$\Psi = C \cos S, \tag{12}$$

where S is a rapidly varying phase and C is a slowly varying amplitude. The dominant terms in the Wheeler–DeWitt equation then are

$$[-(\partial S/\partial y)^2 + (\partial S/\partial x)^2 + V]\,\Psi = 0. \tag{13}$$

This is the same as the classical Hamilton–Jacobi equation obtained from the classical hamiltonian by replacing the momenta P_x and P_y by $\partial S/\partial x$ and $\mathrm{d}S/\mathrm{d}y$ respectively. Thus the integral curves of ∇S correspond to the trajectories of *lorentzian* solutions of the classical field equations.

The wave function shown in fig. 2 does not oscillate in the region under the curve $V = 0$. Thus the classical solutions that correspond to it by the WKB approximation must have a minimum radius or "bounce" at or near $V = 0$. At the minimum radius $\mathrm{d}\phi/\mathrm{d}t$ will be small but $|\phi|$ can be large. Figs. 3a and 3b show the graphs of a and ϕ against t for a classical solution with these initial conditions. It starts off with a period in which a is exponentially increasing and ϕ is steadily decreasing. Then ϕ passes through zero and starts oscillating while a expands like a matter dominated universe, reaches a maximum expansion and re-contracts to another bounce. The oscillatory part of the wave function can be regarded as a superposition of quantum states that are peaked around a family of different classical solutions like that given in fig. 3. These solutions may be characterized by an infinite sequence of integers which give the number of oscillations of ϕ

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

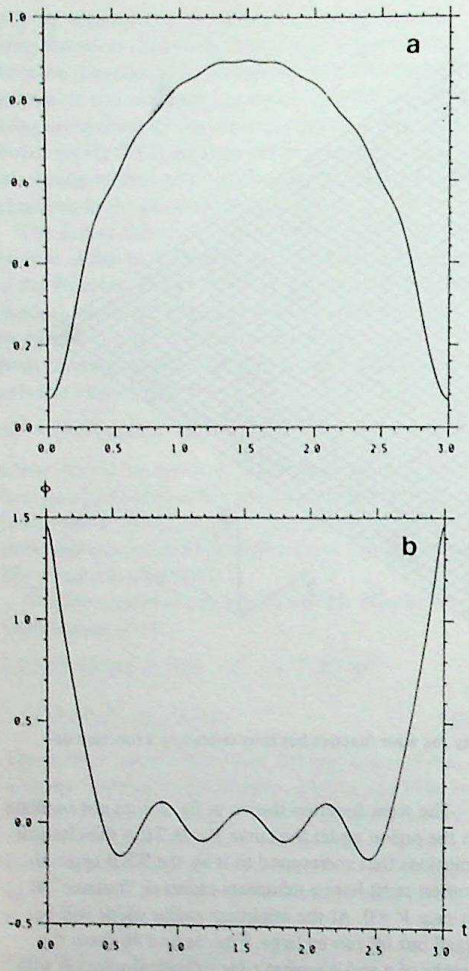


Fig. 3. Graphs of a (a) and ϕ (b) against time for a lorentzian solution of the classical field equations.

between successive minima of a [9]. The solutions that expand out to the present density of the universe will have at least 10^{60} oscillations of ϕ. It would be quite impossible to continue the numerical calculations this far, quite apart from the problem that Ψ contains an exponentially growing component which can be interpreted as corresponding to a positive definite four-geometry in the classical limit.

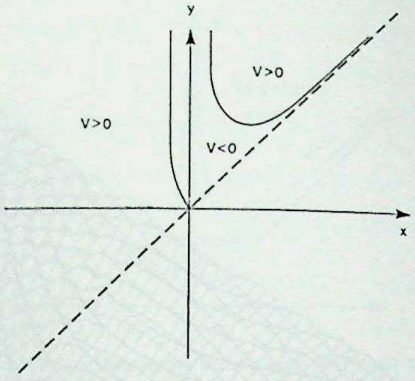


Fig. 4. The x, y plane for the higher derivative model.

The other minisuperspace model we considered had no matter fields but the action contained curvature squared quantum corrections:

$$\hat{I} = -\frac{1}{16\pi}\int \mathrm{d}^4x\,\sqrt{g}\,[R - \alpha C^{\mu\nu\rho\sigma}C_{\mu\nu\rho\sigma} + \beta R^2] + \text{surface term}, \tag{14}$$

here $C_{\mu\nu\rho\sigma}$ is the Weyl tensor. A homogeneous isotropic universe with this action has the hamiltonian [6]:

$$H = \tfrac{1}{4}P_x^2 - \tfrac{1}{4}P_y^2 + V(x,y), \tag{15}$$

where

$$x = 2\beta a R, \quad y = 2a(1+\beta R),$$

$$V = -\tfrac{1}{4}(y^2 - x^2) + (x^2/72\pi\beta)(x-y)^2.$$

The classical equations of the model can be derived from the hamiltonian in the normal way and the quantum mechanical Schrödinger equation can be obtained by the normal substitutions $P_x \to \mathrm{i}\partial/\partial x$ and $P_y \to \mathrm{i}\partial/\partial y$:

$$[\partial^2/\partial y^2 - \partial^2/\partial x^2 + 4V(x,y)]\,\Psi = 0. \tag{16}$$

One has to solve eq. (16) on the half-space $y > x$ (fig. 4) which corresponds to $a > 0$. The action for euclidean solutions to points near the null line $y = x$ ($a = 0$) is very small. Thus the boundary condition on $y = x$ is $\Psi = 1$. The line $y = x$ is not a Cauchy surface by itself: it has to be supplemented by a knowledge of Ψ and its time derivative on a "spacelike" line through

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

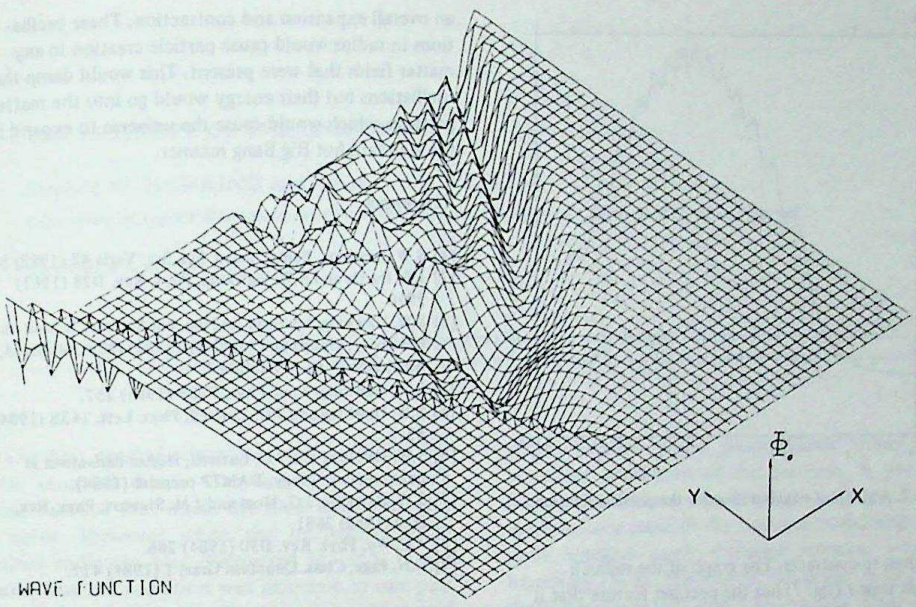


Fig. 5. Numerical results for Ψ in the region $y > |x|$. The wave function Ψ has been divided by a similar factor and has been set to 1 in the region $y < |x|$.

the origin to the region of negative x or by a knowledge of Ψ on the other null line through the origin, $y = -x$. The semi-classical approximation to the path integral (4) indicates that Ψ decreases exponentially on spacelike lines from the origin. We therefore assumed that on the null line $y = -x$, Ψ has the form

$$\Psi = \exp(-a^4/36\pi\beta). \tag{17}$$

Fig. 5 shows the results of the numerical integration for $\beta = 0.001/9\pi$. As predicted from the semi-classical approximation [6], the wave function starts to oscillate near the right-hand branch of the curve $V = 0$ in fig. 4. This behaviour is very similar to that of the massive scalar field model. Again the oscillating part of the wave function can be interpreted via the WKB approximation as corresponding to a family of solutions of the lorentzian field equations. The trajectory in the (x, y) plane of one such solution is shown in fig. 6 for $\beta = 0.2/9\pi$. Again it starts off with an exponential expansion phase and then goes over to a matter dominated era, reaches a maximum expansion

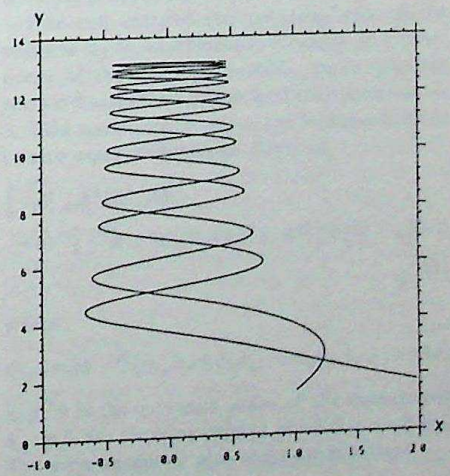


Fig. 6. The trajectory in the x, y plane of a lorentzian solution of the classical field equations.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

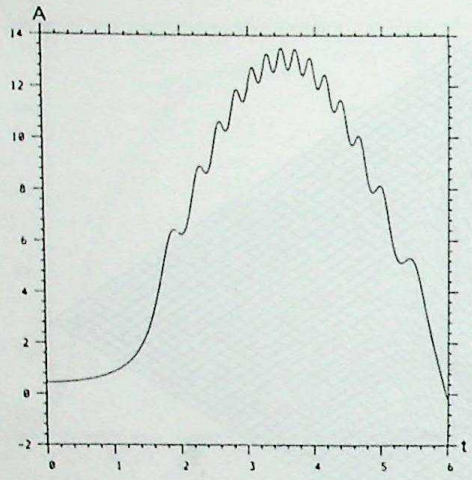


Fig. 7. A graph of a against time for the classical solution.

and then re-contracts. The graph of the radius a against time t (fig. 7) has the peculiar feature that it has small amplitude rapid oscillations superimposed on an overall expansion and contraction. These oscillations in radius would cause particle creation in any matter fields that were present. This would damp the oscillations but their energy would go into the matter particles which would cause the universe to expand in the standard hot Big Bang manner.

References

[1] S.W. Hawking, Pontif. Acad. Sci. Scr. Varia 48 (1982) 563.
[2] J.B. Hartle and S.W. Hawking, Phys. Rev. D28 (1983) 2960.
[3] S.W. Hawking, Lectures les Houches Summer School on Relativity, groups and topology (1983) (North-Holland, Amsterdam), to be published.
[4] S.W. Hawking, Nucl. Phys. B239 (1984) 257.
[5] S.W. Hawking and J.C. Luttrell, Phys. Lett. 143B (1984) 83.
[6] S.W. Hawking and J.C. Luttrell, Higher derivatives in quantum cosmology, DAMTP preprint (1984).
[7] S.W. Hawking, I.G. Moss and J.M. Stewart, Phys. Rev. D10 (1982) 2681.
[8] Z.C. Wu, Phys. Rev. D30 (1984) 286.
[9] D.N. Page, Class. Quantum Grav. 1 (1984) 417.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

THE ISOTROPY OF THE UNIVERSE

Stephen W. HAWKING and Julian C. LUTTRELL

Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, England

Received 26 April 1984

It is shown that the anisotropy of a spatially homogeneous anisotropic minisuperspace model of the universe is likely to be very small at the present time if the universe is in the quantum state defined by a path integral over compact four-geometries.

The first general relativistic models of the universe assumed that it was isotropic about each point because that made the field equations easy to solve. However when the microwave background radiation was discovered in 1965, it was found that the radiation was isotropic to one part in 10^4 apart from a dipole anisotropy which was attributed to the peculiar motion of our galaxy. Since 1965 there have been a number of attempts to show why the universe should be so isotropic. One possible explanation [1] involved the anthropic principle. However this is not to the taste of many people. Another class of explanation involved dissipative processes in the early universe such as neutrino viscosity [2] or particle creation [3]. Such effects would certainly reduce any anisotropy that was present initially. However they failed as explanations of the present isotropy of the universe because one could always make a cosmological model with all arbitrary amount of anisotropy at the present time and simply run one model back in time to find out what initial state it must have started from (ref. [4]). A similar objection applies to the inflationary model of the universe [5]. A period of exponential expansion would certainly reduce an initial anisotropy but if the initial anisotropy is high enough the universe would never get into an exponentially expanding phase [6].

The above attempts to explain the isotropy of the universe failed because they did not specify the initial conditions of the universe. One cannot deduce anything about the present state of the universe without some assumptions about the boundary conditions of the universe. A possible boundary condition that has been proposed is that the quantum state of the universe is defined by a path integral over compact metrics without boundary [7–9]. The aim of this paper is to show that this boundary condition would lead to a quantum state that would correspond in the classical limit to solutions which had a long period of exponential expansion and which were very nearly isotropic at the present time.

One can describe the quantum state of the universe by a wavefunction Ψ which is a functional of the three geometry h_{ij} on a spacelike surface S and of the matter field configurations on S. This wavefunction obeys the Wheeler–DeWitt or zero energy Schrödinger Equation

$$\{-G_{ijkl}\delta^2/\delta h_{ij}\delta h_{kl} + h^{1/2}[-{}^3R + 16\pi T_{00}(\delta/\delta\phi, \phi)]\}\Psi[h_{ij}, \phi] = 0, \quad (1)$$

where

$$G_{ijkl} = \tfrac{1}{2}h^{-1/2}(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}), \quad (2)$$

and ${}^3R$ is the curvature scalar of the three-metric h_{ij} on S. We are using units in which $G = c = \hbar = 1$. The wavefunction Ψ also obeys the equations

$$(\delta\Psi/\delta h_{ij})_{|i} = 8\pi T^{0j}, \quad (3)$$

where a vertical stroke denotes covariant differ-

0370-2693/84/$03.00 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

entiation in S. This equation expresses the invariance of the wavefunction Ψ under coordinate transformations in S.

A solution of the Wheeler–DeWitt equation (1) and the constraint equations (3) describes a possible quantum state of the universe. The particular solution Ψ that corresponds to the quantum state specified by a path integral over compact four-geometries is

$$\Psi[h_{ij},\phi] = \int \mathrm{d}[g_{\mu\nu}]\mathrm{d}[\phi]\exp\{-\hat{I}[g_{\mu\nu},\phi]\}, \quad (4)$$

where $\hat{I}$ is the euclidean action

$$\hat{I} = \frac{1}{16\pi}\int \mathrm{d}^4x\,\sqrt{g}\,R - \frac{1}{8\pi}\int \mathrm{d}^3x\,\sqrt{h}\,K - \int \mathrm{d}^4x\,\sqrt{g}\,L_{\mathrm{m}} \quad (5)$$

and the path integral is taken over all compact four-geometries which have the three-manifold S as their boundary on which the induced metric is h_{ij} and the matter field configuration is ϕ. It can be verified that the wave-function defined by (4) automatically satisfies eqs. (1) and (3) [8–10].

The Wheeler–DeWitt equation is a functional differential equation on superspace, the space of all three-metrics h_{ij} and matter field configurations ϕ on S. It is difficult to solve such an equation on an infinite dimensional space but one can get some idea of the nature of the solution by approximating superspace by a finite dimensional space. In other words one restricts the infinite number of degrees of freedom of the gravitational and matter fields down to a finite number and then solves the Wheeler–DeWitt equation on a finite dimensional "minisuperspace". The simplest example of a minisuperspace is that which corresponds to a homogeneous, isotropic closed universe. In this case there is only one gravitational degree of freedom, the radius a of the three-spheres of homogeneity. In refs. [8] and [9], the wavefunction corresponding to the boundary conditions of compact metrics was found for such a model with a massive scalar field ϕ which was constant on the surfaces of homogeneity. This wavefunction could be interpreted as a superposition of quantum states that, in the classical limit, were peaked around a family of solutions of the classical field equations for a universe filled with a massive scalar field. Most of such solutions have singularities but there is a class without singularities which represents oscillating universes [11,12]. It is members of this class which correspond to the wavefunction given by the path integral (4) over compact geometries.

In this paper we extend the minisuperspace to a homogeneous, anisotropic universe with a scalar field ϕ which is constant on the surfaces of homogeneity. The metric of such a space can be written in the form

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{e}^{2\alpha(t)}[\mathrm{e}^{2\beta}]_{ij}\omega^i\omega^j, \quad (6)$$

where $\beta_{ij}(t)$ is a 3×3 symmetric tracefree matrix and ω^i are three one-forms on the three-sphere which obey

$$\mathrm{d}\omega^i = \epsilon_{ijk}\omega^j \wedge \omega^k. \quad (7)$$

Without loss of generality we can take β to be diagonal and can parametrise it by

$$\beta_{ij} = \mathrm{diag}(\beta_+ + \sqrt{3}\beta_-,\ \beta_+ - \sqrt{3}\beta_-,\ -2\beta_+). \quad (8)$$

The minisuperspace for this model is four dimensional, parametrised by the coordinates (α, β_+, β_-, ϕ). The Wheeler–DeWitt equation is

$$\left[\mathrm{e}^{-p\alpha}(\partial/\partial\alpha)\,\mathrm{e}^{p\alpha}\partial/\partial\alpha - \partial^2/\partial\phi^2 - \partial^2/\partial\beta_+^2 - \partial^2/\partial\beta_-^2 + \mathrm{e}^{6\alpha}m^2\phi^2 + \mathrm{e}^{4\alpha}V(\beta_+,\beta_-)\right]\Psi = 0, \quad (9)$$

where

$$V(\beta_+,\beta_-) = \tfrac{1}{2}\left[\exp(-8\beta_+) + \exp(4\beta_+ + 4\sqrt{3}\beta_-) + \exp(4\beta_+ - 4\sqrt{3}\beta_-)\right] - \left[\exp(4\beta_+) + \exp(-2\beta_+ + 2\sqrt{3}\beta_-) + \exp(-2\beta_+ - 2\sqrt{3}\beta_-)\right] \quad (10)$$

and p represents some, but not all of the ambiguity in the factor ordering, or, equivalently, of the measure in the path integral. Its value will not affect our conclusions. Eq. (9) can be regarded as a wave equation in the flat four-dimensional space with coordinates (α, β_+, β_-, ϕ) with α playing the role of time coordinate. The equation has a varia-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ble M^2 given by

$$M^2 = e^{4\alpha} V(\beta_+, \beta_-) + e^{6\alpha} m^2 \phi^2. \quad (11)$$

As in the case of the isotropic universe [8,9] we have to find a solution of eq. (9) with the boundary conditions that at small three-geometries i.e. at large negative α, the wavefunction should be given by the semi-classical approximation to the path integral (4) over compact four-geometries.

$$\Psi = A\, e^{-B}, \quad (12)$$

where B is the action of a solution of the classical field equations for positive definitive compact metrics which are bounded by a three-sphere with the given values of α, β_+, β_-, and ϕ. If α is sufficiently negative i.e. the three-sphere is small enough, the effect of the scalar field can be neglected. One can then show that there is always a real positive-definite compact four-metric which is a solution of the Einstein equations and which is bounded by the three-sphere with the given α, β_+ and β_-. As α tends to minus infinity along a line of constant β_+ and β_-, the action B of this solution will tend to zero. On the other hand, as α tends to minus infinity along a null geodesic in the α, β_+, β_- plane. B will tend to plus infinity. The behaviour of the prefactor A in eq. (12) will depend on the factor ordering or, equivalently, on the measure in the path integral but in general it will be some polynomial in α, β_+ and β_-.

The wavefunction at large negative α has an exponential or power law behaviour corresponding to a real euclidean four geometry in the classical limit. However, as α increases, the wavefunction will start to oscillate if M^2 is positive and has a gradient which is "timelike" in the α, β_+, β_-, ϕ space. The oscillations will effectively start on the set Q which is the portion of the surface $M^2 = 1$ which is spacelike. An oscillating wavefunction can be interpreted in terms of lorentzian four-geometries [8,9].

One can make the WKB ansatz

$$\Psi = C \cos S, \quad (13)$$

where S is a rapidly varying phase factor and C is a slowly varying phase amplitude. By the Hamilton–Jacobi equation, the integral curves of ∇S in the $(\alpha, \beta_+, \beta_-, \phi)$ space correspond to the trajectories of solutions of the classical lorentzian equations for a homogeneous, anisotropic universe filled with a massive scalar field ϕ. The trajectories of ∇S are well-defined only to the future of the set Q and are approximately orthogonal to Q. This defines the initial conditions for a family of lorentzian solutions of the classical field equations. The quantum state of the universe can be regarded as a superposition of quantum states peaked around different members of this family. If the lorentzian solutions are continued back in time they would in general have a singularity of the "Mixmaster" type in which the anisotropy would oscillate an infinite number of times [13,14]. However because Ψ does not oscillate at values of α more negative than Q, the WKB approximation breaks down in this region and the early stages of the lorentzian solutions do not correspond to the wavefunction. The concept of time becomes fuzzy when the universe is very small.

The requirement that the surface $M^2 = 1$ be spacelike implies that the second term in eq. (11) is dominant i.e.

$$m^2\phi^2 > e^{-2\alpha} V(\beta_+, \beta_-) \quad (14)$$

The requirement that the trajectories be approximately orthogonal to $M^2 = 1$ implies that $d\beta_+/dt$, $d\beta_-/dt$ and $d\phi/dt$ are small. Lorentzian solutions with these initial conditions will expand exponentially at first because the term $m^2\phi^2$ in the action will act like a cosmological constant. During the exponential expansion β_i, will be approximately constant. The magnitude of the scalar field ϕ will decrease but if the initial value $|\phi_i|$ is large compared to one i.e. to the Planck mass, the scale factor e^α will increase by order $\exp(\frac{3}{2}\phi_i^2)$ before ϕ passes through zero. The scalar field will then oscillate in time with a period $2\pi m^{-1}$. It will behave like pressure free matter and will cause the universe to expand with e^α proportional to $t^{2/3}$. If ϕ was coupled to other fields with zero or low rest mass, it would decay into particles of these fields and the universe would expand with e^α proportional to $t^{1/2}$. The universe would expand to some exponentially large radius and then re-collapse.

During the period of exponential expansion, the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

scalar field energy density

$$\rho_m = (d\phi/dt)^2 + m^2\phi^2 \quad (15)$$

will remain constant while the "anisotropy energy density"

$$\rho_\beta = (d\beta_+/dt)^2 + (d\beta_-/dt)^2 + e^{-2\alpha}(V + \tfrac{3}{2}) \quad (16)$$

will decrease like $e^{-2\alpha}$. At the end of the exponential period ρ_m will therefore be greater than ρ_β by a factor of order $\exp(9\phi_1^2)$. During the subsequent expansion of the universe by a factor e^α of order $10^{32}\, m^{1/2}$ up to the present day, the matter energy density will decrease like $e^{-3\alpha}$ or $e^{-4\alpha}$ according to whether the universe is matter or radiation dominated while the anisotropy energy density will decrease as $e^{-2\alpha}$ or faster.

Thus, at the present day

$$\rho_\beta/\rho_m < 10^{61}\, m \exp(-9\phi_1^2). \quad (17)$$

In order for the period of exponential expansion to be long enough for the universe to expand to its present density, $\exp(9\phi_1^2)$ has to be greater than $10^{61}\, m$ i.e.

$$|\phi_1| > 7.5 + \tfrac{1}{2}\ln m. \quad (18)$$

The value of $|\Psi|^2$ at the beginning of the exponential expansion will be of the order of 1 except in the region of finite area in the β_+, β_- plane in which $V(\beta) < 0$ where it will be of the order of $\exp(2/3m^2\phi_1^2)$. Given therefore that inequality (18) is satisfied, it is highly probable that

$$|\phi_1| > 9 + \tfrac{1}{2}\ln m. \quad (19)$$

This would ensure that the present value of ρ_β/ρ_m would be less than the observed upper limit of 10^{-11}. It would also imply that the present density was very near the critical value.

Thus it seems that the proposed boundary condition that the quantum state of the universe is determined by a path integral over compact metrics can account for the observed isotropy of the universe with a high degree of probability. It seems likely that it might also account for the large scale homogeneity with small fluctuations that developed into galaxies.

The authors are very grateful to Don Page for pointing out an error in a preliminary version of this paper.

References

[1] C.B. Collins and S.W. Hawking, Astrophys. J. 180 (1973) 317.
[2] C.W. Misner, Astrophys. J. 151 (1968) 431.
[3] Y.B. Zeldovich and A.A. Starobinsky, Zh. Eksp. Teor. Fiz. 61 (1972) 2161 [Sov. Phys. JETP. 34 (1972) 1159].
[4] J.M. Stewart, Mon. Not. R. Astron. Soc. 145 (1969) 347.
[5] A.H. Guth, Phys. Rev. D23 (1981) 347.
[5] A.D. Linde, Phys. Lett. 108B (1982) 389.
[6] J.D. Barrow, Quart. J.R. Astron. Soc. 23 (1982) 344.
[7] S.W. Hawking, Pontificae Academiae Scientarum, Scripta Varia 48 (1982) 563.
[8] S.W. Hawking, Lectures Les Houches Summer School on relativity and topology (1983) (North-Holland, Amsterdam) to be published.
[9] S.W. Hawking, Nucl. Phys. B239 (1984) 257.
[10] J.B. Hartle and S.W. Hawking Phys. Rev. D. 28 (1983) 2960.
[11] D.N. Page, A fractal set of perpetually bouncing universes Pennsylvania State University preprint.
[12] A.D. Linde, Phys. Lett. 129B (1983) 117.
[13] C.W. Misner, Phys. Rev. Lett. 22 (1969) 1071.
[14] V.A. Belinskii, E.M. Lifschitz and I.M. Khalatnikov, Usp. Fiz. Nauk. 102 (1970) 463 [Sov. Phys. Usp. 13 (1970) 475]; Adv. Phys. 19 (1970) 525.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PHYSICAL REVIEW D

PARTICLES AND FIELDS

THIRD SERIES, VOLUME 31, NUMBER 8 15 APRIL 1985

Origin of structure in the Universe

J. J. Halliwell and S. W. Hawking

Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, United Kingdom and Max Planck Institut for Physics and Astrophysics, Foehringer Ring 6, Munich, Federal Republic of Germany

(Received 17 December 1984)

It is assumed that the Universe is in the quantum state defined by a path integral over compact four-metrics. This can be regarded as a boundary condition for the wave function of the Universe on superspace, the space of all three-metrics and matter field configurations on a three-surface. We extend previous work on finite-dimensional approximations to superspace to the full infinite-dimensional space. We treat the two homogeneous and isotropic degrees of freedom exactly and the others to second order. We justify this approximation by showing that the inhomogeneous or anisotropic modes start off in their ground state. We derive time-dependent Schrödinger equations for each mode. The modes remain in their ground state until their wavelength exceeds the horizon size in the period of exponential expansion. The ground-state fluctuations are then amplified by the subsequent expansion and the modes reenter the horizon in the matter- or radiation-dominated era in a highly excited state. We obtain a scale-free spectrum of density perturbations which could account for the origin of galaxies and all other structure in the Universe. The fluctuations would be compatible with observations of the microwave background if the mass of the scalar field that drives the inflation is 10^{14} GeV or less.

I. INTRODUCTION

Observations of the microwave background indicate that the Universe is very close to homogeneity and isotropy on a large scale. Yet we know that the early Universe cannot have been completely homogeneous and isotropic because in that case galaxies and stars would not have formed. In the standard hot big-bang model the density perturbations required to produce these structures have to be assumed as initial conditions. However, in the inflationary model of the Universe[1–4] it was possible to show that the ground-state fluctuations of the scalar field that causes the exponential expansion would lead to a spectrum of density perturbations that was almost scale free.[5–7] In the simplest grand-unified-theory (GUT) inflationary model the amplitude of the density perturbations was too large but an amplitude that was consistent with observation could be obtained in other models with a different potential for the scalar field.[8] Similarly, ground-state fluctuations of the gravitational-wave modes would lead to a spectrum of long-wavelength gravitational waves that would be consistent with observation provided that the Hubble constant H in the inflationary period was not more than about 10^{-4} of the Planck mass.[9]

One cannot regard these results as a completely satisfactory explanation of the origin of structure in the Universe because the inflationary model does not make any assumption about the initial or boundary conditions of the Universe. In particular, it does not guarantee that there should be a period of exponential expansion in which the scalar field and the gravitational-wave modes would be in the ground state. In the absence of some assumption about the boundary conditions of the Universe, any present state would be possible: one could pick an arbitrary state for the Universe at the present time and evolve it backward in time to see what initial conditions it arose from. It has recently been proposed[10–13] that the boundary conditions of the Universe are that it has no boundary. In other words, the quantum state of the Universe is defined by a path integral over compact four-metrics without boundary. The quantum state can be described by a wave function Ψ which is a function on the infinite-dimensional space $\mathcal{W}$ called superspace which consists of all three-metrics h_{ij} and matter field configurations Φ_0 on a three-surface S. Because the wave function does not depend on time explicitly, it obeys a system of zero-energy Schrödinger equations, one for each choice of the shift N_i and the lapse N on S. The Schrödinger equations can be decomposed into the momentum constraints, which imply that the wave function is the same at all points of $\mathcal{W}$ that are related by coordinate transformations, and the Wheeler-DeWitt equations, which can be

 ©1985 The American Physical Society

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

regarded as a system of second-order differential equations for Ψ on W. The requirement that the wave function be given by a path integral over compact four-metrics then becomes a set of boundary conditions for the Wheeler-DeWitt equations which determines a unique solution for Ψ.

It is difficult to solve differential equations on an infinite-dimensional manifold. Attention has therefore been concentrated on finite-dimensional approximations to W, called "minisuperspaces." In other words, one restricts the number of gravitational and matter degrees of freedom to a finite number and then solves the Wheeler-DeWitt equations on a finite-dimensional manifold with boundary conditions that reflect the fact that the wave function is given by a path integral over compact four-metrics. In particular,[12–15] it has been shown that in the case of a homogeneous isotropic closed universe of radius a with a massive scalar field ϕ the wave function corresponds in the classical limit to a family of classical solutions which have a long period of exponential or "inflationary" expansion and then go over to a matter-dominated expansion, reach a maximum radius, and then collapse in a time-symmetric manner. This model would be in agreement with observation but, because it is so restricted, the only prediction it can make is that the observed value of the density parameter Ω should be exactly one.[15] The aim of this paper is to extend this minisuperspace model to the full number of degrees of freedom of the gravitational and scalar fields. We treat the 2 degrees of freedom of the minisuperspace model exactly and we expand the other inhomogeneous and anisotropic degrees of freedom to second order in the Hamiltonian. In the region of W in which Ψ oscillates rapidly, one can use the WKB approximation to relate the wave function to a family of classical solutions and so introduce a concept of time. As in the minisuperspace case, the family includes solutions with a long period of exponential expansion. We show that the gravitational-wave and density-perturbation modes obey decoupled time-dependent Schrödinger equations with respect to the time parameter of the classical solution. The boundary conditions imply that these modes start off in the ground state. While they remain within the horizon of the exponentially expanding phase, they can relax adiabatically and so they remain in the ground state. However, when they expand outside the horizon of the inflationary period, they become "frozen" until they reenter the horizon in the matter-dominated era. They then give rise to gravitational waves and a scale-free spectrum of density perturbations. These would be consistent with the observations of the microwave background and could be large enough to explain the origins of galaxies if the mass of the scalar field were about 10^{-5} of the Planck mass. Thus the proposal that the quantum state of the Universe is defined by a path integral over compact four-metrics seems to be able to account for the origin of structure in the Universe: it arises, not from arbitrary initial conditions, but from the ground-state fluctuations that have to be present by the Heisenberg uncertainty principle.

In Sec. II we review the Hamiltonian formalism of classical general relativity, and in Sec. III we show how this leads to the canonical treatment of the quantum theory. In Sec. IV we summarize earlier work[13] on a homogeneous isotropic minisuperspace model with a massive scalar field. We extend this to all the matter and gravitational degrees of freedom in Sec. V, treating the inhomogeneous modes to second order in the Hamiltonian. In Sec. VI we decompose the wave function into a background term which obeys an equation similar to that of the unperturbed minisuperspace model, and perturbation terms which obey time-dependent Schrödinger equations. We use the path-integral expression for the wave function in Sec. VII to show that the perturbation wave functions start out in their ground states. Their subsequent evolution is described in Sec. VIII. In Sec. IX we calculate the anisotropy that these perturbations would produce in the microwave background and compare with observation. In Sec. X we summarize the paper and conclude that the proposed quantum state could account not only for the large-scale homogeneity and isotropy but also for the structure on smaller scales.

II. CANONICAL FORMULATION OF GENERAL RELATIVITY

We consider a compact three-surface S which divides the four-manifold M into two parts. In a neighborhood of S one can introduce a coordinate t such that S is the surface $t=0$ and coordinates x^i $(i=1,2,3)$. The metric takes the form

$$ds^2=-(N^2-N_iN^i)dt^2+2N_idx^idt+h_{ij}dx^idx^j\ . \tag{2.1}$$

N is called the lapse function. It measure the proper-time separation of surfaces of constant t. N_i is called the shift vector. It measures the deviation of the lines of constant x^i from the normal to the surface S. The action is

$$I=\int (L_g+L_m)d^3x\,dt\ , \tag{2.2}$$

where

$$L_g=\frac{m_P{}^2}{16\pi}N(G^{ijkl}K_{ij}K_{kl}+h^{1/2}{}^3R)\ , \tag{2.3}$$

$$K_{ij}=\frac{1}{2N}\left[-\frac{\partial h_{ij}}{\partial t}+2N_{(i\,|j)}\right]\ , \tag{2.4}$$

is the second fundamental form of S, and

$$G^{ijkl}=\tfrac{1}{2}h^{1/2}(h^{ik}h^{jl}+h^{il}h^{jk}-2h^{ij}h^{kl})\ . \tag{2.5}$$

In the case of a massive scalar field Φ

$$L_m=\tfrac{1}{2}Nh^{1/2}\left[N^{-2}\left[\frac{\partial\Phi}{\partial t}\right]^2-2\frac{N^i}{N^2}\frac{\partial\Phi}{\partial t}\frac{\partial\Phi}{\partial x^i}-\left[h^{ij}-\frac{N^iN^j}{N^2}\right]\frac{\partial\Phi\partial\Phi}{\partial x^i\partial x^j}-m^2\Phi^2\right]\ . \tag{2.6}$$

In the Hamiltonian treatment of general relativity one regards the components h_{ij} of the three-metric and the field Φ as the canonical coordinates. The canonically conjugate momenta are

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$\pi^{ij}=\frac{\partial L_g}{\partial \dot{h}_{ij}}=-\frac{h^{1/2}m_P{}^2}{16\pi}(K^{ij}-h^{ij}K)\,, \tag{2.7}$$

$$\pi_\Phi=\frac{\partial L_m}{\partial \dot{\Phi}}=N^{-1}h^{1/2}\left[\dot{\Phi}-N^i\frac{\partial \Phi}{\partial x^i}\right]. \tag{2.8}$$

The Hamiltonian is

$$\begin{aligned}H&=\int(\pi^{ij}\dot{h}_{ij}+\pi_\Phi\dot{\Phi}-L_g-L_m)d^3x\\&=\int(NH_0+N_iH^i)d^3x\,,\end{aligned} \tag{2.9}$$

where

$$H_0=16\pi m_P{}^{-2}G_{ijkl}\pi^{ij}\pi^{kl}-\frac{m_P{}^2}{16\pi}h^{1/2}{}^3R+\tfrac{1}{2}h^{1/2}\left[\frac{\pi_\Phi{}^2}{h}+h^{ij}\frac{\partial\Phi\partial\Phi}{\partial x^i\partial x^j}+m^2\Phi^2\right], \tag{2.10}$$

$$H^i=-2\pi^{ij}{}_{|j}+h^{ij}\frac{\partial\Phi}{\partial x^j}\pi_\Phi\,, \tag{2.11}$$

and

$$G_{ijkl}=\tfrac{1}{2}h^{-1/2}(h_{ik}h_{jl}+h_{il}h_{jk}-h_{ij}h_{kl})\,. \tag{2.12}$$

The quantities N and N_i are regarded as Lagrange multipliers. Thus the solution obeys the momentum constraint

$$H^i=0 \tag{2.13}$$

and the Hamiltonian constraint

$$H_0=0\,. \tag{2.14}$$

For given fields N and N^i on S the equations of motion are

$$\dot{h}_{ij}=\frac{\partial H}{\partial\pi^{ij}},\quad \dot{\pi}^{ij}=-\frac{\partial H}{\partial h_{ij}}\,,$$
$$\dot{\Phi}=\frac{\partial H}{\partial\pi_\Phi},\quad \dot{\pi}_\Phi=-\frac{\partial H}{\partial\Phi}\,. \tag{2.15}$$

III. QUANTIZATION

The quantum state of the Universe can be described by a wave function Ψ which is a function on the infinite-dimensional manifold W of all three-metrics h_{ij} and matter fields Φ on S. A tangent vector to W is a pair of fields (γ_{ij},μ) on S where γ_{ij} can be regarded as a small change of the metric h_{ij} and μ can be regarded as a small change of Φ. For each choice of $N>0$ on S there is a natural metric $\Gamma(N)$ on W:[15]

$$ds^2=\int N^{-1}\left[\frac{m_P{}^2}{32\pi}G^{ijkl}\gamma_{ij}\gamma_{kl}+\tfrac{1}{2}h^{1/2}\mu^2\right]d^3x\,. \tag{3.1}$$

The wave function Ψ does not depend explicitly on the time t because t is just a coordinate which can be given arbitrary values by different choices of the undetermined multipliers N and N_i. This means that Ψ obeys the zero-energy Schrödinger equation:

$$H\Psi=0\,. \tag{3.2}$$

The Hamiltonian operator H is the classical Hamiltonian with the usual substitutions:

$$\pi^{ij}(x)\to -i\frac{\delta}{\delta h_{ij}(x)},\quad \pi_\phi(x)\to -i\frac{\delta}{\delta\phi(x)}\,. \tag{3.3}$$

Because N and N_i are regarded as independent Lagrange multipliers, the Schrödinger equation can be decomposed into two parts. There is the momentum constraint

$$\begin{aligned}H_-\Psi&\equiv\int N_iH^id^3x\ \Psi\\&=\int h^{1/2}N_i\left[2\left\{\frac{\delta}{\delta h_{ij}(x)}\right\}_{|j}-h^{ij}\frac{\partial\Phi}{\partial x^j}\frac{\delta}{\delta\Phi(x)}\right]d^3x\ \Psi\\&=0\,.\end{aligned} \tag{3.4}$$

This implies that Ψ is the same on three-metrics and matter field configurations that are related by coordinate transformations in S. The other part of the Schrödinger equation, corresponding to $H_|\Psi=0$, where $H_|=\int NH_0d^3x$ is called the Wheeler-DeWitt equation. There is one Wheeler-DeWitt equation for each choice of N on S. One can regard them as a system of second-order partial differential equations for Ψ on W. There is some ambiguity in the choice of operator ordering in these equations but this will not affect the results of this paper. We shall assume that $H_|$ has the form[15]

$$(-\tfrac{1}{2}\nabla^2+\xi R+V)\Psi=0\,, \tag{3.5}$$

where ∇^2 is the Laplacian in the metric $\Gamma(N)$. R is the curvature scalar of this metric and the potential V is

$$V=\int h^{1/2}N\left[-\frac{m_P{}^2}{16\pi}{}^3R+\epsilon+U\right]d^3x\,, \tag{3.6}$$

where $U=T^{00}-\frac{1}{2}\pi_\Phi{}^2$. The constant ϵ can be regarded as a renormalization of the cosmological constant Λ. We shall assume that the renormalized Λ is zero. We shall also assume that the coefficient ξ of the scalar curvature R of W is zero.

Any wave function Ψ which satisfies the momentum constraint and the Wheeler-DeWitt equation for each choice of N and N_i on S describes a possible quantum state of the Universe. We shall be concerned with the particular solution which represents the quantum state defined by a path integral over compact four-metrics without boundary. In this case[11–13]

$$\Psi=\int d[g_{\mu\nu}]d[\Phi]\exp[-\hat{I}(g_{\mu\nu},\Phi)]\,, \tag{3.7}$$

where $\hat{I}$ is the Euclidean action obtained by setting N negative imaginary and the path integral is taken over all compact four-metrics $g_{\mu\nu}$ and matter fields Φ which are bounded by S on which the three-metric is h_{ij} and the matter field is Φ. One can regard (3.7) as a boundary condition on the Wheeler-DeWitt equations. It implies that Ψ tends to a constant, which can be normalized to one, as h_{ij} goes to zero.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

IV. UNPERTURBED FRIEDMANN MODEL

References 12–14 considered the minisuperspace model which consisted of a Friedmann model with metric

$$ds^2=\sigma^2(-N^2dt^2+a^2d\Omega_3^2)\,, \tag{4.1}$$

where $d\Omega_3^2$ is the metric of the unit three-sphere. The normalization factor $\sigma^2=2/3\pi m_P^2$ has been included for convenience. The model contains a scalar field $(2^{1/2}\pi\sigma)^{-1}\phi$ with mass $\sigma^{-1}m$ which is constant on surfaces of constant t. One can easily generalize this to the case of a scalar field with a potential $V(\phi)$. Such generalizations include models with higher-derivative quantum corrections.[16] The action is

$$I=-\tfrac{1}{2}\int dt\,Na^3\left[\frac{1}{N^2a^2}\left[\frac{da}{dt}\right]^2-\frac{1}{a^2}-\frac{1}{N^2}\left[\frac{d\phi}{dt}\right]^2+m^2\phi^2\right]. \tag{4.2}$$

The classical Hamiltonian is

$$H=\tfrac{1}{2}N(-a^{-1}\pi_a^2+a^{-3}\pi_\phi^2-a+a^3m^2\phi^2)\,, \tag{4.3}$$

where

$$\pi_a=-\frac{a\,da}{N\,dt},\quad \pi_\phi=\frac{a^3d\phi}{N\,dt}\,. \tag{4.4}$$

The classical Hamiltonian constraint is $H=0$. The classical field equations are

$$N\frac{d}{dt}\left[\frac{1}{N}\frac{d\phi}{dt}\right]+\frac{3}{a}\frac{da}{dt}\frac{d\phi}{dt}+N^2m^2\phi=0\,, \tag{4.5}$$

$$N\frac{d}{dt}\left[\frac{1}{N}\frac{da}{dt}\right]=N^2am^2\phi^2-2a\left[\frac{d\phi}{dt}\right]^2. \tag{4.6}$$

The Wheeler-DeWitt equation is

$$\tfrac{1}{2}Ne^{-3\alpha}\left[\frac{\partial^2}{\partial\alpha^2}-\frac{\partial^2}{\partial\phi^2}+2V\right]\Psi(\alpha,\phi)=0\,, \tag{4.7}$$

where

$$V=\tfrac{1}{2}(e^{6\alpha}m^2\phi^2-e^{4\alpha}) \tag{4.8}$$

and $\alpha=\ln a$. One can regard Eq. (4.7) as a hyperbolic equation for Ψ in the flat space with coordinates (α,ϕ) with α as the time coordinate. The boundary condition that gives the quantum state defined by a path integral over compact four-metrics is $\Psi\to1$ as $\alpha\to-\infty$. If one integrates Eq. (4.7) with this boundary condition, one finds that the wave function starts oscillating in the region $V>0$, $|\phi|>1$ (this has been confirmed numerically[14]). One can interpret the oscillatory component of the wave function by the WKB approximation:

$$\Psi=\mathrm{Re}(Ce^{iS})\,, \tag{4.9}$$

where C is a slowly varying amplitude and S is a rapidly varying phase. One chooses S to satisfy the classical Hamilton-Jacobi equation:

$$H(\pi_\alpha,\pi_\phi,\alpha,\phi)=0\,, \tag{4.10}$$

where

$$\pi_\alpha=\frac{\partial S}{\partial\alpha},\quad \pi_\phi=\frac{\partial S}{\partial\phi}\,. \tag{4.11}$$

One can write (4.10) in the form

$$\tfrac{1}{2}f^{ab}\frac{\partial S\partial S}{\partial q^a\partial q^b}+e^{-3\alpha}V=0\,, \tag{4.12}$$

where f^{ab} is the inverse to the metric $\Gamma(1)$:

$$f^{ab}=e^{-3\alpha}\mathrm{diag}(-1,1)\,. \tag{4.13}$$

The wave function (4.9) will then satisfy the Wheeler-DeWitt equation if

$$\nabla^2C+2if^{ab}\frac{\partial C\partial S}{\partial q^a\partial q^b}+iC\nabla^2S=0\,, \tag{4.14}$$

where ∇^2 is the Laplacian in the metric f_{ab}. One can ignore the first term in Eq. (4.14) and can integrate the equation along the trajectories of the vector field $X^a=dq^a/dt=f^{ab}\partial S/\partial q^b$ and so determine the amplitude C. These trajectories correspond to classical solutions of the field equations. They are parametrized by the coordinate time t of the classical solutions.

The solutions that correspond to the oscillating part of the wave function of the minisuperspace model start out at $V=0$, $|\phi|>1$ with $d\alpha/dt=d\phi/dt=0$. They expand exponentially with

$$S=-\tfrac{1}{3}e^{3\alpha}m\,|\phi|(1-m^{-2}e^{-2\alpha}\phi^{-2})$$

$$\approx-\tfrac{1}{3}e^{3\alpha}m\,|\phi|\,, \tag{4.15}$$

$$\frac{d\alpha}{dt}=m\,|\phi|,\quad \frac{d\,|\phi|}{dt}=-\tfrac{1}{3}m\,. \tag{4.16}$$

After a time of order $3m^{-1}(|\phi_1|-1)$, where ϕ_1 is the initial value of ϕ, the field ϕ starts to oscillate with frequency m. The solution then becomes matter dominated and expands with e^α proportional to $t^{2/3}$. If there were other fields present, the massive scalar particles would decay into light particles and then the solution would expand with e^α proportional to $t^{1/2}$. Eventually the solution would reach a maximum radius of order $\exp(9\phi_1^2/2)$ or $\exp(9\phi_1^2)$ depending on whether it is radiation or matter dominated for most of the expansion. The solution would then recollapse in a similar manner.

V. THE PERTURBED FRIEDMANN MODEL

We assume that the metric is of the form (2.1) except the right hand side has been multiplied by a normalization factor σ^2. The three-metric h_{ij} has the form

$$h_{ij}=a^2(\Omega_{ij}+\epsilon_{ij})\,, \tag{5.1}$$

where Ω_{ij} is the metric on the unit three-sphere and ϵ_{ij} is a perturbation on this metric and may be expanded in harmonics:

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$\epsilon_{ij}=\sum_{n,l,m}[6^{1/2}a_{nlm}\tfrac{1}{3}\Omega_{ij}Q^n_{lm}+6^{1/2}b_{nlm}(P_{ij})^n_{lm}+2^{1/2}c^0_{nlm}(S^0_{ij})^n_{lm}+2^{1/2}c^e_{nlm}(S^e_{ij})^n_{lm}+2\,d^0_{nlm}(G^0_{ij})^n_{lm}+2\,d^e_{nlm}(G^e_{ij})^n_{lm}]\,. \tag{5.2}$$

The coefficients $a_{nlm},b_{nlm},c^0_{nlm},c^e_{nlm},d^0_{nlm},d^e_{nlm}$ are functions of the time coordinate t but not the three spatial coordinates x^i.

The $Q(x^i)$ are the standard scalar harmonics on the three-sphere. The $P_{ij}(x^i)$ are given by (suppressing all but the i,j indices)

$$P_{ij}=\frac{1}{(n^2-1)}Q_{|ij}+\tfrac{1}{3}\Omega_{ij}Q\,. \tag{5.3}$$

They are traceless, $P_i{}^i=0$. The S_{ij} are defined by

$$S_{ij}=S_{i|j}+S_{j|i}\,, \tag{5.4}$$

where S_i are the transverse vector harmonics, $S_i{}^{|i}=0$. The G_{ij} are the transverse traceless tensor harmonics $G_i{}^i=G_{ij}{}^{|j}=0$. Further details about the harmonics and their normalization can be found in Appendix A.

The lapse, shift, and the scalar field $\Phi(x^i,t)$ can be expanded in terms of harmonics:

$$N=N_0\left[1+6^{-1/2}\sum_{n,l,m}g_{nlm}Q^n_{lm}\right], \tag{5.5}$$

$$N_i=e^{\alpha}\sum_{n,l,m}[6^{-1/2}k_{nlm}(P_i)^n_{lm}+2^{1/2}j_{nlm}(S_i)^n_{lm}]\,, \tag{5.6}$$

$$\Phi=\sigma^{-1}\left[\frac{1}{2^{1/2}\pi}\phi(t)+\sum_{n,l,m}f_{nlm}Q^n_{lm}\right], \tag{5.7}$$

where $P_i=[1/(n^2-1)]Q_{|i}$. Hereafter, the labels n, l, m, o, and e will be denoted simply by n. One can then expand the action to all orders in terms of the "background" quantities a,ϕ,N_0 but only to second order in the "perturbations" $a_n,b_n,c_n,d_n,f_n,g_n,k_n,j_n$:

$$I=I_0(a,\phi,N_0)+\sum_n I_n\,, \tag{5.8}$$

where I_0 is the action of the unperturbed model (4.2) and I_n is quadratic in the perturbations and is given in Appendix B.

One can define conjugate momenta in the usual manner. They are

$$\pi_\alpha=-N_0{}^{-1}e^{3\alpha}\dot\alpha+\text{quadratic terms}\,, \tag{5.9}$$

$$\pi_\phi=N_0{}^{-1}e^{3\alpha}\dot\phi+\text{quadratic terms}\,, \tag{5.10}$$

$$\pi_{a_n}=-N_0{}^{-1}e^{3\alpha}[\dot a_n+\dot\alpha(a_n-g_n)+\tfrac{1}{3}e^{-\alpha}k_n]\,, \tag{5.11}$$

$$\pi_{b_n}=N_0{}^{-1}e^{3\alpha}\frac{(n^2-4)}{(n^2-1)}(\dot b_n+4\dot\alpha b_n-\tfrac{1}{3}e^{-\alpha}k_n)\,, \tag{5.12}$$

$$\pi_{c_n}=N_0{}^{-1}e^{3\alpha}(n^2-4)(\dot c_n+4\dot\alpha c_n-e^{-\alpha}j_n)\,, \tag{5.13}$$

$$\pi_{d_n}=N_0{}^{-1}e^{3\alpha}(\dot d_n+4\dot\alpha d_n)\,, \tag{5.14}$$

$$\pi_{f_n}=N_0{}^{-1}e^{3\alpha}[\dot f_n+\dot\phi(3a_n-g_n)]\,. \tag{5.15}$$

The quadratic terms in Eqs. (5.9) and (5.10) are given in Appendix B. The Hamiltonian can then be expressed in terms of these momenta and the other quantities:

$$H=N_0\left[H_{|0}+\sum_n H^n_{|2}+\sum_n g_nH^n_{|1}\right]+\sum_n(k_n{}^SH^n_{-1}+j_n{}^VH^n_{-1})\,. \tag{5.16}$$

The subscripts 0,1,2 on the $H_|$ and H_- denote the orders of the quantities in the perturbations and S and V denote the scalar and vector parts of the shift part of the Hamiltonian. $H_{|0}$ is the Hamiltonian of the unperturbed model with $N=1$:

$$H_{|0}=\tfrac{1}{2}e^{-3\alpha}(-\pi_\alpha{}^2+\pi_\phi{}^2+e^{6\alpha}m^2\phi^2-e^{4\alpha})\,. \tag{5.17}$$

The second-order Hamiltonian is given by

$$H_{|2}=\sum_n H^n_{|2}=\sum_n({}^SH^n_{|2}+{}^VH^n_{|2}+{}^TH^n_{|2})\,,$$

where

$$\begin{aligned}{}^SH^n_{|2}=\tfrac{1}{2}e^{-3\alpha}\Bigg[&\left[\tfrac{1}{2}a_n{}^2+\frac{10(n^2-4)}{(n^2-1)}b_n{}^2\right]\pi_\alpha{}^2+\left[\tfrac{15}{2}a_n{}^2+\frac{6(n^2-4)}{(n^2-1)}b_n{}^2\right]\pi_\phi{}^2\\&-\pi_{a_n}{}^2+\frac{(n^2-1)}{(n^2-4)}\pi_{b_n}{}^2+\pi_{f_n}{}^2+2a_n\pi_{a_n}\pi_\alpha+8b_n\pi_{b_n}\pi_\alpha-6a_n\pi_{f_n}\pi_\phi\\&-e^{4\alpha}\left[\tfrac{1}{3}(n^2-\tfrac{5}{2})a_n{}^2+\frac{(n^2-7)}{3}\frac{(n^2-4)}{(n^2-1)}b_n{}^2+\tfrac{1}{3}(n^2-4)a_nb_n-(n^2-1)f_n{}^2\right]\\&+e^{6\alpha}m^2(f_n{}^2+6a_nf_n\phi)+e^{6\alpha}m^2\phi^2\left[\tfrac{3}{2}a_n{}^2-\frac{6(n^2-4)}{(n^2-1)}b_n{}^2\right]\Bigg]\,,\end{aligned} \tag{5.18}$$

$${}^VH^n_{|2}=\tfrac{1}{2}e^{-3\alpha}\left[(n^2-4)c_n{}^2(10\pi_\alpha{}^2+6\pi_\phi{}^2)+\frac{1}{(n^2-4)}\pi_{c_n}{}^2+8c_n\pi_{c_n}\pi_\alpha+(n^2-4)c_n{}^2(2e^{4\alpha}-6e^{6\alpha}m^2\phi^2)\right], \tag{5.19}$$

$${}^TH^n_{|2}=\tfrac{1}{2}e^{-3\alpha}\{d_n{}^2(10\pi_\alpha{}^2+6\pi_\phi{}^2)+\pi_{d_n}{}^2+8d_n\pi_{d_n}\pi_\alpha+d_n{}^2[(n^2+1)e^{4\alpha}-6e^{6\alpha}m^2\phi^2]\}\,. \tag{5.20}$$

The first-order Hamiltonians are

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$H^n_{\uparrow 1}=\tfrac{1}{2}e^{-3\alpha}\{-a_n(\pi_\alpha{}^2+3\pi_\phi{}^2)+2(\pi_\phi\pi_{f_n}-\pi_\alpha\pi_{a_n})+m^2e^{6\alpha}(2f_n\phi+3a_n\phi^2)-\tfrac{2}{3}e^{4\alpha}[(n^2-4)b_n+(n^2+\tfrac{1}{2})a_n]\}\,. \tag{5.21}$$

The shift parts of the Hamiltonian are

$${}^SH^n_{\downarrow 1}=\tfrac{1}{3}e^{-3\alpha}\left[-\pi_{a_n}+\pi_{b_n}+\left[a_n+\frac{4(n^2-4)}{(n^2-1)}b_n\right]\pi_\alpha+3f_n\pi_\phi\right]\,, \tag{5.22}$$

$${}^VH^n_{\downarrow 1}=e^{-\alpha}[\pi_{c_n}+4(n^2-4)c_n\pi_\alpha]\,. \tag{5.23}$$

The classical field equations are given in Appendix B.

Because the Lagrange multipliers N_0,g_n,k_n,j_n are independent, the zero energy Schrödinger equation

$$H\Psi=0 \tag{5.24}$$

can be decomposed as before into momentum constraints and Wheeler-DeWitt equations. As the momentum constraints are linear in the momenta, there is no ambiguity in the operator ordering. One therefore has

$${}^SH^n_{\downarrow 1}\Psi=-\tfrac{1}{3}e^{-3\alpha}\left[\frac{\partial}{\partial a_n}-\left[a_n+\frac{4(n^2-4)}{(n^2-1)}b_n\right]\frac{\partial}{\partial\alpha}-\frac{\partial}{\partial b_n}-3f_n\frac{\partial}{\partial\phi}\right]\Psi=0\,, \tag{5.25}$$

$${}^VH^n_{\downarrow 1}\Psi=e^{-\alpha}\left[\frac{\partial}{\partial c_n}+4(n^2-4)c_n\frac{\partial}{\partial\alpha}\right]\Psi=0\,. \tag{5.26}$$

The first-order Hamiltonians $H^n_{\uparrow 1}$ give a series of finite dimensional second-order differential equations, one for each n. In the order of approximation that we are using, the ambiguity in the operator ordering will consist of the possible addition of terms linear in $\partial/\partial\alpha$. The effect of such terms can be compensated for by multiplying the wave function by powers of e^α. This will not affect the relative probabilities of different observations at a given value of α. We shall therefore ignore such ambiguities and terms:

$$\tfrac{1}{2}e^{-3\alpha}\left[a_n\left(\frac{\partial^2}{\partial\alpha^2}+3\frac{\partial^2}{\partial\phi^2}\right)-2\left(\frac{\partial^2}{\partial f_n\partial\phi}-\frac{\partial^2}{\partial a_n\partial\alpha}\right)+m^2e^{6\alpha}[2\phi f_n+3a_n\phi^2]-\tfrac{2}{3}e^{4\alpha}[(n^2-4)b_n+(n^2+\tfrac{1}{2})a_n]\right]\Psi=0\,. \tag{5.27}$$

Finally, one has an infinite-dimensional second-order differential equation

$$\left[H_{\uparrow 0}+\sum_n({}^SH^n_{\uparrow 2}+{}^VH^n_{\uparrow 2}+{}^TH^n_{\uparrow 2})\right]\Psi=0\,, \tag{5.28}$$

where $H_{\uparrow 0}$ is the operator in the Wheeler-DeWitt equation of the unperturbed Friedmann minisuperspace model:

$$H_{\uparrow 0}=\tfrac{1}{2}e^{-3\alpha}\left[\frac{\partial^2}{\partial\alpha^2}-\frac{\partial^2}{\partial\phi^2}+e^{6\alpha}m^2\phi^2-e^{4\alpha}\right] \tag{5.29}$$

and

$$\begin{aligned}{}^SH^n_{\uparrow 2}=\tfrac{1}{2}e^{-3\alpha}\Bigg[&-\left[\tfrac{1}{2}a_n{}^2+\frac{10(n^2-4)}{(n^2-1)}b_n{}^2\right]\frac{\partial^2}{\partial\alpha^2}-\left[\tfrac{15}{2}a_n{}^2+\frac{6(n^2-4)}{(n^2-1)}b_n{}^2\right]\frac{\partial^2}{\partial\phi^2}\\&+\frac{\partial^2}{\partial a_n{}^2}-\frac{(n^2-1)}{(n^2-4)}\frac{\partial^2}{\partial b_n{}^2}-\frac{\partial^2}{\partial f_n{}^2}-2a_n\frac{\partial^2}{\partial a_n\partial\alpha}-8b_n\frac{\partial^2}{\partial b_n\partial\alpha}+6a_n\frac{\partial^2}{\partial f_n\partial\phi}\\&-e^{4\alpha}\left[\tfrac{1}{3}(n^2-\tfrac{5}{2})a_n{}^2+\frac{(n^2-7)}{3}\frac{(n^2-4)}{(n^2-1)}b_n{}^2+\tfrac{2}{3}(n^2-4)a_nb_n-(n^2-1)f_n{}^2\right]\\&+e^{6\alpha}m^2(f_n{}^2+6a_nf_n\phi)+e^{6\alpha}m^2\phi^2\left[\tfrac{3}{2}a_n{}^2-\frac{6(n^2-4)}{(n^2-1)}b_n{}^2\right]\Bigg]\,,\end{aligned} \tag{5.30}$$

$${}^VH^n_{\uparrow 2}=\tfrac{1}{2}e^{-3\alpha}\left[-(n^2-4)c_n{}^2\left[10\frac{\partial^2}{\partial\alpha^2}+6\frac{\partial^2}{\partial\phi^2}\right]-\frac{1}{(n^2-4)}\frac{\partial^2}{\partial c_n{}^2}-8c_n\frac{\partial^2}{\partial c_n\partial\alpha}+(n^2-4)c_n{}^2(2e^{4\alpha}-6e^{6\alpha}m^2\phi^2)\right]\,, \tag{5.31}$$

$${}^TH^n_{\uparrow 2}=\tfrac{1}{2}e^{-3\alpha}\left[-d_n{}^2\left[10\frac{\partial^2}{\partial\alpha^2}+6\frac{\partial^2}{\partial\phi^2}\right]-\frac{\partial^2}{\partial d_n{}^2}-8d_n\frac{\partial^2}{\partial d_n\partial\alpha}+d_n{}^2[(n^2+1)e^{4\alpha}-6e^{6\alpha}m^2\phi^2]\right]\,. \tag{5.32}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

We shall call Eq. (5.28) the master equation. It is not hyperbolic because, as well as the positive second derivatives $\partial^2/\partial\alpha^2$ in $H_{\restriction 0}$, there are the positive second derivatives $\partial^2/\partial a_n{}^2$ in each ${}^S H^n_{\restriction 2}$. However, one can use the momentum constraint (5.25) to substitute for the partial derivatives with respect to a_n and then solve the resultant differential equation on $a_n=0$. Similarly, one can use the momentum constraint (5.26) to substitute for the partial derivatives with respect to c_n and then solve on $c_n=0$. One thus obtains a modified equation which is hyperbolic for small f_n. If one knows the wave function on $a_n=0=c_n$, one can use the momentum constraints to calculate the wave function at other values of a_n and c_n.

VI. THE WAVE FUNCTION

Because the perturbation modes are not coupled to each other, the wave function can be expressed as a sum of terms of the form

$$\Psi=\mathrm{Re}\left[\Psi_0(\alpha,\phi)\prod_n \Psi^{(n)}(\alpha,\phi,a_n,b_n,c_n,d_n,f_n)\right]$$

$$=\mathrm{Re}(C\,e^{iS})\,, \tag{6.1}$$

where S is a rapidly varying function of α and ϕ and C is a slowly varying function of all the variables. If one substitutes (6.1) into the master equation and divides by Ψ, one obtains

$$-\frac{\nabla_2{}^2\Psi_0}{2\Psi_0}-\sum_n\frac{\nabla_2{}^2\Psi^{(n)}}{2\Psi^{(n)}}-\sum_{n\lesssim m}\frac{(\nabla_2\Psi^{(n)})\cdot(\nabla_2\Psi^{(m)})}{2\Psi^{(n)}\Psi^{(m)}}$$

$$-\frac{(\nabla_2\Psi_0)}{\Psi_0}\cdot\left[\sum_n\frac{\nabla_2\Psi^{(n)}}{\Psi^{(n)}}\right]$$

$$+\sum_n\frac{H^n_{\restriction 2}\Psi}{\Psi}+e^{-3\alpha}V(\alpha,\phi)=0\,, \tag{6.2}$$

where $\nabla_2{}^2$ is the Laplacian in the minisuperspace metric $f_{ab}=e^{3\alpha}\mathrm{diag}(-1,1)$ and the dot product is with respect to this metric.

An individual perturbation mode does not contribute a significant fraction of the sums in the third and fourth terms in Eq. (6.2). Thus these terms can be replaced by

$$-\frac{(\nabla_2\Psi)}{\Psi}\cdot\sum_n\frac{(\nabla_2\Psi^{(n)})}{\Psi^{(n)}}+\frac{1}{2}\left[\sum_n\frac{\nabla_2\Psi^{(n)}}{\Psi^{(n)}}\right]^2$$

$$\approx -i(\nabla_2 S)\cdot\sum_n\frac{(\nabla_2\Psi^{(n)})}{\Psi^{(n)}}+\frac{1}{2}\left[\sum_n\frac{\nabla_2\Psi^{(n)}}{\Psi^{(n)}}\right]^2. \tag{6.3}$$

In order that the ansatz (6.1) be valid, the terms in (6.2) that depend on a_n,b_n,c_n,d_n,f_n have to cancel out. This implies

$$\frac{(\nabla_2\Psi)}{\Psi}\cdot(\nabla_2\Psi^{(n)})+\tfrac{1}{2}\nabla_2{}^2\Psi^{(n)}=\frac{H^n_{\restriction 2}\Psi}{\Psi}\Psi^{(n)}\,, \tag{6.4}$$

$$(-\tfrac{1}{2}\nabla_2{}^2+e^{-3\alpha}V+\tfrac{1}{2}J\cdot J)\Psi_0=0\,, \tag{6.5}$$

where

$$J=\sum_n\frac{\nabla_2\Psi^{(n)}}{\Psi^{(n)}}\,.$$

In regions in which the phase S is a rapidly varying function of α and ϕ, one can neglect the second term in (6.4) in comparison with the first term. One can also replace the π_α and π_ϕ which appear in $H^n_{\restriction 2}$ by $\partial S/\partial\alpha$ and $\partial S/\partial\phi$, respectively. The vector $X^a=f^{ab}\partial S/\partial q^b$ obtained by raising the covector $\nabla_2 S$ by the inverse minisuperspace metric f^{ab} can be regarded as $\partial/\partial t$ where t is the time parameter of the classical Friedmann metric that corresponds to Ψ by the WKB approximation. One then obtains a time dependent Schrödinger equation for each mode along a trajectory of the vector field X^a:

$$i\frac{\partial\Psi^{(n)}}{\partial t}=H^n_{\restriction 2}\Psi^{(n)}\,. \tag{6.6}$$

Equation (6.5) can be interpreted as the Wheeler-DeWitt equation for a two-dimensional minisuperspace model with an extra term $\frac{1}{2}J\cdot J$ arising from the perturbations. In order to make J finite, one will have to make subtractions. Subtracting out the ground-state energies of the $H^n_{\restriction 2}$ corresponds to a renormalization of the cosmological constant Λ. There is a second subtraction which corresponds to a renormalization of the Planck mass m_P and a third one which corresponds to a curvature-squared counterterm. The effect of such higher-derivative terms in the action has been considered elsewhere.[16]

One can write $\Psi^{(n)}$ as

$$\Psi^{(n)}={}^S\Psi^{(n)}(\alpha,\phi,a_n,b_n f_n)\,{}^V\Psi^{(n)}(\alpha,\phi,c_n)\,{}^T\Psi^{(n)}(\alpha,\phi,d_n)\,, \tag{6.7}$$

where ${}^S\Psi^{(n)}$, ${}^V\Psi^{(n)}$, and ${}^T\Psi^{(n)}$ obey independent Schrödinger equations with ${}^S H^n_{\restriction 2}$, ${}^V H^n_{\restriction 2}$, and ${}^T H^n_{\restriction 2}$, respectively.

VII. THE BOUNDARY CONDITIONS

We want to find the solution of the master equation that corresponds to

$$\Psi[h_{ij},\Phi]=\int d[g_{\mu\nu}]d[\Phi]\exp(-\hat{I})\,, \tag{7.1}$$

where the integral is taken over all compact four-metrics and matter fields which are bounded by the three-surface S. If one takes the scale parameter α to be very negative but keeps the other parameters fixed, the Euclidean action $\hat{I}$ tends to zero like $e^{2\alpha}$. Thus one would expect Ψ to tend to one as α tends to minus infinity.

One can estimate the form of the scalar, vector, and tensor parts ${}^S\Psi^{(n)}$, ${}^V\Psi^{(n)}$, ${}^T\Psi^{(n)}$ of the perturbation $\Psi^{(n)}$ from the path integral (7.1) One takes the four-metric $g_{\mu\nu}$ and the scalar field Φ to be of the background form

$$ds^2=\sigma^2(-N^2dt^2+e^{2\alpha(t)}d\Omega_3{}^2) \tag{7.2}$$

and $\phi(t)$, respectively, plus a small perturbation described by the variables (a_n,b_n,f_n), c_n, and d_n as functions of t. In order for the background four-metric to be compact, it has to be Euclidean when $\alpha=-\infty$, i.e., N has to be purely negative imaginary at $\alpha=-\infty$, which we shall take to be $t=0$. In regions in which the metric is Lorentzian, N

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

will be real and positive. In order to allow a smooth transition from Euclidean to Lorentzian, we shall take N to be of the form $-ie^{i\mu}$ where $\mu=0$ at $t=0$. In order that the four-metric and the scalar field be regular at $t=0, a_n, b_n, c_n, d_n, f_n$ have to vanish there.

The tensor perturbations d_n have the Euclidean action

$$^{T}\hat{I}_n=\tfrac{1}{2}\int dt\, d_n\, {}^{T}D d_n + \text{boundary term}\,, \tag{7.3}$$

where

$$^{T}D=\left[-\frac{d}{dt}\left[\frac{e^{3\alpha}d}{iN_0dt}\right]+iN_0e^{\alpha}(n^2-1)\right]+4iN_0e^{3\alpha}\left[+\tfrac{1}{2}e^{-2\alpha}-\tfrac{1}{2}m^2\phi^2-\frac{3\dot{\phi}^2}{2(iN_0)^2}-\frac{3\dot{\alpha}^2}{2(iN_0)^2}-\frac{1d}{iN_0dt}\left[\frac{\dot{\alpha}}{iN_0}\right]\right]. \tag{7.4}$$

The last term in (7.4) vanishes if the background metric satisfies the background field equations. The action is extremized when d_n satisfies the equation

$$^{T}Dd_n=0\,. \tag{7.5}$$

For a d_n that satisfies (7.5), the action is just the boundary term

$$^{T}\hat{I}_n^{\rm cl}=\frac{1}{2iN_0}e^{3\alpha}(d_n\dot{d}_n+4\dot{\alpha}d_n^{\,2})\,. \tag{7.6}$$

The path integral over d_n will be

$$\int d[d_n]\exp(-{}^{T}\hat{I}_n)=(\det{}^{T}D)^{-1/2}\exp(-{}^{T}\hat{I}_n^{\rm cl})\,. \tag{7.7}$$

One now has to integrate (7.7) over different background metrics to obtain the wave function $^{T}\Psi^{(n)}$. One expects the dominant contribution to come from background metrics that are near a solution of the classical background field equations. For such metrics one can employ the adiabatic approximation in which one regards α to be a slowly varying function of t. Then the solution of (7.5) which obeys the boundary condition $d_n=0$ at $t=0$ is

$$d_n=A(e^{\nu\tau}-e^{-\nu\tau})\,, \tag{7.8}$$

where $\nu=e^{-\alpha}(n^2-1)^{1/2}$ and $\tau=\int iN_0dt$. This approximation will be valid for background fields which are near a solution of the background field equations and for which

$$\left|\frac{\dot{\alpha}}{N_0}\right|\ll n\,e^{-\alpha}\,. \tag{7.9}$$

For a regular Euclidean metric, $|\dot{\alpha}/N_0|=e^{-\alpha}$ near $t=0$. If the metric is a Euclidean solution of the background field equations, then $|\dot{\alpha}/N_0|<e^{-\alpha}$. Thus the adiabatic approximation should hold for large values of n into the region in which the solution of the background field equations becomes Lorentzian and the WKB approximation can be used. The wave function $^{T}\Psi^{(n)}$ will then be

$$^{T}\Psi^{(n)}=B\exp\left[-\left[\tfrac{1}{2}n\,e^{2\alpha}\coth(\nu\tau)+\frac{2}{iN_0}\dot{\alpha}\,e^{3\alpha}\right]d_n^{\,2}\right]. \tag{7.10}$$

In the Euclidean region, τ will be real and positive. For large values of n, $\coth(\nu\tau)\approx 1$. In the Lorentzian region where the WKB approximation applies, τ will be complex but it will still have a positive real part and $\coth(\nu\tau)$ will still be approximately 1 for large n. Thus

$$^{T}\Psi^{(n)}=B\exp\left[-2i\frac{\partial S}{\partial\alpha}d_n^{\,2}-\tfrac{1}{2}n\,e^{2\alpha}d_n^{\,2}\right]. \tag{7.11}$$

The normalization constant B can be chosen to be 1. Thus, apart from a phase factor, the gravitational-wave modes enter the WKB region in their ground state.

We now consider the vector part $^{V}\Psi^{(n)}$ of the wave function. This is pure gauge as the quantities c_n can be given any value by gauge transformations parametrized by the j_n. The freedom to make gauge transformations is reflected quantum mechanically in the constraint

$$e^{-\alpha}\left[\frac{\partial}{\partial c_n}+4(n^2-4)c_n\frac{\partial}{\partial\alpha}\right]\Psi=0\,. \tag{7.12}$$

One can integrate (7.12) to give

$$\Psi(\alpha,\{c_n\})=\Psi\left[\alpha-2\sum_n(n^2-4)c_n^{\,2},0\right], \tag{7.13}$$

where the dependence on the other variables has been suppressed. One can also replace $\partial\Psi/\partial\alpha$ by $i(\partial S/\partial\alpha)\Psi$. One can then solve for $^{V}\Psi^{(n)}$:

$$^{V}\Psi^{(n)}=\exp\left[2i(n^2-4)c_n^{\,2}\frac{\partial S}{\partial\alpha}\right]. \tag{7.14}$$

The scalar perturbation modes a_n, b_n, and f_n involve a combination of the behavior of the tensor and vector perturbations. The scalar part of the action is given in Appendix B. The action is extremized by solutions of the classical equations

$$N_0\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{a}_n}{N_0}\right]+\tfrac{1}{3}(n^2-4)N_0^{\,2}e^{\alpha}(a_n+b_n)+3e^{3\alpha}(\dot{\phi}\dot{f}_n-N_0^{\,2}m^2\phi f_n)$$

$$=N_0^{\,2}[3e^{3\alpha}m^2\phi^2-\tfrac{1}{3}(n^2+2)e^{\alpha}]g_n+e^{3\alpha}\dot{\alpha}\dot{g}_n-\tfrac{1}{3}N_0\frac{d}{dt}\left[e^{2\alpha}\frac{k_n}{N_0}\right] \tag{7.15}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$N_0\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{b}_n}{N_0}\right]-\tfrac{1}{3}(n^2-1)N_0{}^2e^{\alpha}(a_n+b_n)=\tfrac{1}{3}(n^2-1)N_0{}^2e^{\alpha}g_n+\tfrac{1}{3}N_0\frac{d}{dt}\left[e^{2\alpha}\frac{k_n}{N_0}\right], \tag{7.16}$$

$$N_0\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{f}_n}{N_0}\right]+3e^{3\alpha}\dot{\phi}\dot{a}_n+N_0{}^2[m^2e^{3\alpha}+(n^2-1)e^{\alpha}]f_n=e^{3\alpha}(-2N_0{}^2m^2\phi g_n+\dot{\phi}\dot{g}_n-e^{-\alpha}\dot{\phi}k_n)\,. \tag{7.17}$$

There is a three-parameter family of solutions to (7.15)–(7.17) which obey the boundary condition $a_n=b_n=f_n=0$ at $t=0$. There are however, two constraint equations:

$$\dot{a}_n+\frac{(n^2-4)}{(n^2-1)}\dot{b}_n+3f_n\dot{\phi}=\dot{\alpha}g_n-\frac{e^{-\alpha}}{(n^2-1)}k_n\,, \tag{7.18}$$

$$3a_n(-\dot{\alpha}^2+\dot{\phi}^2)+2(\dot{\phi}\dot{f}_n-\dot{\alpha}\dot{a}_n)+N_0{}^2m^2(2f_n\phi+3a_n\phi^2)-\tfrac{2}{3}N_0{}^2e^{-2\alpha}[(n^2-4)b_n+(n^2+\tfrac{1}{2})a_n]$$
$$=\tfrac{2}{3}\dot{\alpha}e^{-\alpha}k_n+2g_n(-\dot{\alpha}^2+\dot{\phi}^2)\,. \tag{7.19}$$

These correspond to the two gauge degrees of freedom parametrized by k_n and g_n, respectively. The Euclidean action for a solution to Eqs. (7.15)–(7.19) is

$$S\hat{I}_n^{\rm cl}=\frac{1}{2iN_0}e^{3\alpha}\left[-a_n\dot{a}_n+\frac{(n^2-4)}{(n^2-1)}b_n\dot{b}_n+f_n\dot{f}_n+\dot{\alpha}\left[-a_n{}^2+\frac{4(n^2-4)}{(n^2-1)}b_n{}^2\right]+3\dot{\phi}a_nf_n+g_n(\dot{\alpha}a_n-\dot{\phi}f_n)\right.$$
$$\left.-\tfrac{1}{3}e^{-\alpha}k_n\left[a_n+\frac{(n^2-4)}{(n^2-1)}b_n\right]\right], \tag{7.20}$$

where the background field equations have been used.

In many ways the simplest gauge to work in is that with $g_n=k_n=0$. However, this gauge does not allow one to find a compact four-metric which is bounded by a three-surface with arbitrary values of a_n, b_n, and f_n and which is a solution of the Eqs. (7.15)–(7.17) and the constraint equations. Instead, we shall use the gauge $a_n=b_n=0$ and shall solve the constraint Eqs. (7.18) and (7.19) to find g_n and k_n:

$$g_n=3\frac{(n^2-1)\dot{\alpha}\dot{\phi}f_n+\dot{\phi}\dot{f}_n+N_0{}^2m^2\phi f_n}{(n^2-4)\dot{\alpha}^2+3\dot{\phi}^2}\,, \tag{7.21}$$

$$k_n=3(n^2-1)e^{\alpha}\frac{\dot{\alpha}\dot{\phi}\dot{f}_n+N_0{}^2m^2\phi f_n\dot{\alpha}-3f_n\dot{\phi}(-\dot{\alpha}^2+\dot{\phi}^2)}{(n^2-4)\dot{\alpha}^2+3\dot{\phi}^2}\,. \tag{7.22}$$

With these substituted, (7.17) becomes a second-order equation for f_n,

$$N_0\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{f}_n}{N_0}\right]+N_0{}^2[m^2e^{3\alpha}+(n^2-1)e^{\alpha}]f_n=e^{3\alpha}(-2N_0{}^2m^2\phi g_n+\dot{\phi}\dot{g}_n-e^{-\alpha}\dot{\phi}k_n)\,. \tag{7.23}$$

For large n we can again use the adiabatic approximation to estimate the solution of (7.23) when $|\phi|>1$:

$$f_n=A\sinh(\nu\tau)\,, \tag{7.24}$$

where $\nu^2=e^{-2\alpha}(n^2-1)$. Thus for these modes

$${}^S\Psi^{(n)}(\alpha,\phi,0,0,f_n)\approx\exp\left[-\tfrac{1}{2}n\,e^{2\alpha}f_n{}^2-\tfrac{1}{2}i\frac{\partial S}{\partial\phi}g_nf_n\right]. \tag{7.25}$$

This is of the ground-state form apart from a small phase factor. The value of ${}^S\Psi^{(n)}$ at nonzero values of a_n and b_n can be found by integrating the constraint equations (5.25) and (5.27).

The tensor and scalar modes start off in their ground states, apart possibly from the modes at low n. The vector modes are pure gauge and can be neglected. Thus the total energy

$$E=\sum_n\frac{H^{(n)}_{\frac{1}{2}}\Psi^{(n)}}{\Psi^{(n)}}$$

of the perturbations will be small when the ground-state energies are subtracted. But $E=i(\nabla_2S)\cdot J$ where $J=\sum_n\nabla_2\Psi^{(n)}/\Psi^{(n)}$. Thus J is small. This means that the wave function Ψ_0 will obey the Wheeler-DeWitt equation of the unperturbed minisuperspace model and the phase factor S will be approximately $-i\ln\Psi_0$. However the homogeneous scalar field mode ϕ will not start out in its ground state. There are two reasons for this: first, regularity at $t=0$ requires $a_n=b_n=c_n=d_n=f_n=0$, but

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

does not require $\phi=0$. Second, the classical field equation for ϕ is of the form of a damped harmonic oscillator with a constant frequency m rather than a decreasing frequency $e^{-\alpha}n$. This means that the adiabatic approximation is not valid at small t and that the solution of the classical field equation is ϕ approximately constant. The action of such solutions is small, so large values of $|\phi|$ are not damped as they are for the other variables. Thus the WKB trajectories which start out from large values of $|\phi|$ have high probability. They will correspond to classical solutions which have a long inflationary period and then go over to a matter-dominated expansion. In a realistic model which included other fields of low rest mass, the matter energy in the oscillations of the massive scalar field would decay into light particles with a thermal spectrum. The model would then expand as a radiation-dominated universe.

VIII. GROWTH OF PERTURBATIONS

The tensor modes will obey the Schrödinger equation

$$i\frac{\partial {}^{T}\Psi^{(n)}}{\partial t}={}^{T}H^{n}_{\frac{1}{2}}{}^{T}\Psi^{(n)} \tag{8.1}$$

$$=\tfrac{1}{2}e^{-3\alpha}\left\{+{d_n}^2\left[10\left(\frac{\partial S}{\partial\alpha}\right)^2+6\left(\frac{\partial S}{\partial\phi}\right)^2\right]\right.$$

$$-\frac{\partial^2}{\partial {d_n}^2}-8d_n i\frac{\partial S}{\partial\alpha}\frac{\partial}{\partial d_n}$$

$$\left.+{d_n}^2[(n^2+1)e^{4\alpha}-6e^{6\alpha}m^2\phi^2]\right\}. \tag{8.2}$$

One can write

$${}^{T}\Psi^{(n)}=\exp(-2\alpha)\exp\left[-2i\frac{\partial S}{\partial\alpha}{d_n}^2\right]{}^{T}\Psi_0^{(n)}, \tag{8.3}$$

then

$$i\frac{\partial {}^{T}\Psi_0^{(n)}}{\partial t}=\tfrac{1}{2}e^{-3\alpha}\left[-\frac{\partial^2}{\partial {d_n}^2}+{d_n}^2(n^2-1)e^{4\alpha}\right]{}^{T}\Psi_0^{(n)}. \tag{8.4}$$

The WKB approximation to the background Wheeler-DeWitt equation has been used in deriving (8.4). Then (8.4) has the form of the Schrödinger equation for an oscillator with a time-dependent frequency $\nu=(n^2-1)^{1/2}e^{-\alpha}$. Initially the wave function ${}^{T}\Psi_0^{(n)}$ will be in the ground state (apart from a normalization factor) and the frequency ν will be large compared to $\dot{\alpha}$. In this case one can use the adiabatic approximation to show that ${}^{T}\Psi_0^{(n)}$ remains in the ground state

$${}^{T}\Psi_0^{(n)}\approx\exp(-\tfrac{1}{2}n\,e^{2\alpha}{d_n}^2). \tag{8.5}$$

The adiabatic approximation will break down when $\nu\approx\dot{\alpha}$, i.e., the wave length of the gravitational mode becomes equal to the horizon scale in the inflationary period. The wave function ${}^{T}\Psi_0^{(n)}$ will then freeze

$${}^{T}\Psi_0^{(n)}\approx\exp(-\tfrac{1}{2}n\,e^{2\alpha_*}{d_n}^2), \tag{8.6}$$

where α_* is the value of α at which the mode goes outside the horizon. The wave function ${}^{T}\Psi_0^{(n)}$ will remain of the form (8.6) until the mode reenters the horizon in the matter- or radiation-dominated era at the much greater value α_r of α. One can then apply the adiabatic approximation again to (8.4) but ${}^{T}\Psi_0^{(n)}$ will no longer be in the ground state; it will be a superposition of a number of highly excited states. This is the phenomenon of the amplification of the ground-state fluctuations in the gravitational-wave modes that was discussed in Refs. 9, 17, and 18.

The behavior of the scalar modes is rather similar but their description is more complicated because of the gauge degrees of freedom. In the previous section we evaluated the wave function ${}^{S}\Psi^{(n)}$ on $a_n=b_n=0$ by the path-integral prescription. The ground-state form (in f_n) that we found will be valid until the adiabatic approximation breaks down, i.e., until the wavelength of the mode exceeds the horizon distance during the inflationary period. In order to discuss the subsequent behavior of the wave function. It is convenient to use the first-order Hamiltonian constraint (5.27) to evaluate ${}^{S}\Psi^{(n)}$ on $a_n\neq0, b_n=f_n=0$. One finds that

$${}^{S}\Psi^{(n)}(\alpha,\phi,a_n,0,0)=B\exp[iC{a_n}^2]{}^{S}\Psi_0^{(n)}(\alpha,\phi,a_n). \tag{8.7}$$

The normalization and phase factors B and C depend on α and ϕ but not a_n:

$$C=\frac{1}{2}\left(\frac{\partial S}{\partial\alpha}\right)^{-1}\left[\left(\frac{\partial S}{\partial\alpha}\right)^2-\tfrac{1}{3}(n^2-4)e^{4\alpha}\right]. \tag{8.8}$$

At the time the wavelength of the mode equals the horizon distance during the inflationary period, the wave function ${}^{S}\Psi_0^{(n)}$ has the form

$${}^{S}\Psi_0^{(n)}=\exp(-\tfrac{1}{2}ny_*{}^{-2}e^{2\alpha_*}{a_n}^2), \tag{8.9}$$

where y_* is the value of $y=(\partial S/\partial\alpha)[\partial S/\partial\phi]^{-1}$ when the mode leaves the horizon, $y_*=3\phi_*$. More generally, in the case of a scalar field with a potential $V(\phi)$, $y=6V(\partial V/\partial\phi)^{-1}$.

One can obtain a Schrödinger equation for ${}^{S}\Psi_0^{(n)}$ by putting $b_n=f_n=0$ in the scalar Hamiltonian ${}^{S}H^{n}_{\frac{1}{2}}$ and substituting for $\partial/\partial b_n$ and $\partial/\partial f_n$ from the momentum constraint (5.25) and the first-order Hamiltonian constraint (5.27), respectively. This gives

$$i\frac{\partial {}^{S}\Psi_0^{(n)}}{\partial t}=\tfrac{1}{2}e^{-3\alpha}\left\{-y^2\frac{\partial^2}{\partial {a_n}^2}+e^{4\alpha}(n^2-4)\right.$$

$$\left.\times\left[\frac{1}{y^2}-\tfrac{1}{3}e^{4\alpha}\left(\frac{\partial S}{\partial\alpha}\right)^{-2}\right]{a_n}^2\right\}{}^{S}\Psi_0^{(n)}, \tag{8.10}$$

where terms of order $1/n^2$ have been neglected. The term $e^{4\alpha}[\partial S/\partial\alpha]^{-2}$ will be small compared to $1/y^2$ except near the time of maximum radius of the background solution. The Schrödinger equation for ${}^{S}\Psi_0^{(n)}(a_n)$ is very similar to the equation for ${}^{T}\Psi_0^{(n)}(d_n)$, (8.4), except that the kinetic term is multiplied by a factor y^2 and the potential term is divided by a factor y^2. One would therefore ex-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

pect that for wavelengths within the horizon. ${}^{S}\Psi_0^{(n)}$ would have the ground-state form $\exp(-\frac{1}{2}ny^{-2}e^{2\alpha}a_n^2)$ and this is borne out by (8.9). On the other hand, when the wavelength becomes larger than the horizon, the Schrödinger equation (8.10) indicates that ${}^{T}\Psi_0^{(n)}$ will freeze in the form (8.9) until the mode reenters the horizon in the matter-dominated era. Even if the equation of state of the Universe changes to radiation dominated during the period that the wavelength of the mode is greater than the horizon size, it will still be true that ${}^{S}\Psi_0^{(n)}$ is frozen in the form (8.9). The ground-state fluctuations in the scalar modes will therefore be amplified in a similar manner to the tensor modes. At the time of reentry of the horizon the rms fluctuation in the scalar modes, in the gauge in which $b_n=f_n=0$, will be greater by the factor y_* than the rms fluctuation in the tensor modes of the same wavelength.

IX. COMPARISON WITH OBSERVATION

From a knowledge of ${}^{T}\Psi_0^{(n)}$ and ${}^{S}\Psi_0^{(n)}$ one can calculate the relative probabilities of observing different values of d_n and a_n at a given point on a trajectory of the vector field X^i, i.e., at a given value of α and ϕ in a background metric which is a solution of the classical field equations. In fact, the dependence on ϕ will be unimportant and we shall neglect it. One can then calculate the probabilities of observing different amounts of anisotropy in the microwave background and can compare these predictions with the upper limits set by observation.

The tensor and scalar perturbation modes will be in highly excited states at large values of α. This means that we can treat their development as an ensemble evolving according to the classical equations of motion with initial distributions in d_n and a_n proportional to $|{}^{T}\Psi_0^{(n)}|^2$ and $|{}^{S}\Psi_0^{(n)}|^2$, respectively. The initial distributions in $\dot{d}_n$ and $\dot{a}_n$ will be proportional to $|{}^{T}\Psi_0^{(n)}\pi_{d_n}{}^{T}\Psi_0^{(n)}|$ and $|{}^{S}\Psi_0^{(n)}\pi_{a_n}{}^{S}\Psi_0^{(n)}|$, respectively. In fact, at the time that the modes reenter the horizon, the distributions will be concentrated at $\dot{d}_n=\dot{a}_n=0$.

The surfaces with $b_n=f_n=0$ will be surfaces of constant energy density in the classical solution during the inflationary period. By local conservation of energy, they will remain surfaces of constant energy density in the era after the inflationary period when the energy is dominated by the coherent oscillations of the homogeneous background scalar field ϕ. If the scalar particles decay into light particles and heat up the Universe, the surfaces with $b_n=f_n=0$ will be surfaces of constant temperature. The surface of last scattering of the microwave background will be such a surface with temperature T_s. The microwave radiation can be considered to have propagated freely to us from this surface. Thus the observed temperature will be

$$T_0=\frac{T_s}{1+z}\,, \tag{9.1}$$

where z is the red-shift of the surface of last scattering. Variations in the observed temperature will arise from variations in z in different directions of observation. These are given by

$$1+z=l^{\mu}n_{\mu} \tag{9.2}$$

evaluated at the surface of last scattering where n_μ is the unit normal to the surfaces of constant t in the gauge $g_n=k_n=j_n=0$ and $b_n=f_n=0$ on the surface of last scattering and l^μ is the parallel propagated tangent vector to the null geodesic from the observer normalized by $l^\mu n_\mu=1$ at the present time. One can calculate the evolution of $l^\mu n_\mu$ down the past light cone of the observer:

$$\frac{d}{d\lambda}[l^{\mu}n_{\mu}]=n_{\mu;\nu}l^{\mu}l^{\nu}\,, \tag{9.3}$$

where λ is the affine parameter on the null geodesic. The only nonzero components of $n_{\mu;\nu}$ are

$$n_{i;j}=e^{2\alpha}\left[\dot{\alpha}\Omega_{ij}+\sum_n(\dot{a}_n+\dot{\alpha}a_n)\tfrac{1}{3}\Omega_{ij}Q+\sum_n(\dot{b}_n+\dot{\alpha}b_n)P_{ij}+\sum_n(\dot{d}_n+\dot{\alpha}d_n)G_{ij}\right]. \tag{9.4}$$

In the gauge that we are using, the dominant anisotropic terms in (9.4) on the scale of the horizon, will be those involving $\dot{\alpha}a_n$ and $\dot{\alpha}d_n$. These will give temperature anisotropies of the form

$$\langle(\Delta T/T)^2\rangle\approx\langle a_n^2\rangle \text{ or } \approx\langle d_n^2\rangle\,. \tag{9.5}$$

The number of modes that contribute to anisotropies on the scale of the horizon is of the order of n^3. From the results of the last section

$$\langle a_n^2\rangle=y_*^2n^{-1}e^{-2\alpha_*}\,, \tag{9.6}$$

$$\langle d_n^2\rangle=n^{-1}e^{-2\alpha_*}\,. \tag{9.7}$$

The dominant contribution comes from the scalar modes which give

$$\langle(\Delta T/T)^2\rangle\approx y_*^2n^2e^{-2\alpha_*}\,. \tag{9.8}$$

But $ne^{-\alpha_*}\approx\dot{\alpha}_*$, the value of the Hubble constant at the time that the present horizon size left the horizon during the inflationary period. The observational upper limit of about 10^{-8} on $\langle(\Delta T/T)^2\rangle$ restricts this Hubble constant to be less than about $5\times10^{-5}m_P$ (Ref. 8) which in turn restricts the mass of the scalar field to be less than 10^{14} GeV.

X. CONCLUSION AND SUMMARY

We started from the proposal that the quantum state of the Universe is defined by a path integral over compact four-metrics. This can be regarded as a boundary condition for the Wheeler-DeWitt equation for the wave function of the Universe on the infinite-dimensional manifold, superspace, the space of all three-metrics and matter field configurations on a three-surface S. Previous papers had considered finite-dimensional approximations to superspace and had shown that the boundary condition led to a wave function which could be interpreted as correspond-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ing to a family of classical solutions which were homogeneous and isotropic and which had a period of exponential or inflationary expansion. In the present paper we extended this work to the full superspace without restrictions. We treated the two basic homogeneous and isotropic degrees of freedom exactly and the other degrees of freedom to second order. We justified this approximation by showing that the inhomogeneous or anisotropic modes started out in their ground states.

We derived time-dependent Schrödinger equations for each mode. We showed that they remained in the ground state until their wavelength exceeded the horizon size during the inflationary period. In the subsequent expansion the ground-state fluctuations got frozen until the wavelength reentered the horizon during the radiation- or matter-dominated era. This part of the calculation is similar to earlier work on the development of gravitational waves[9] and density perturbations[5,6] in the inflationary Universe but it has the advantage that the assumptions of a period of exponential expansion and of an initial ground state for the perturbations are justified. The perturbations would be compatible with the upper limits set by observations of the microwave background if the scalar field that drives the inflation has a mass of 10^{14} GeV or less.

In Sec. VIII we calculated the scalar perturbations in a gauge in which the surfaces of constant time are surfaces of constant density. There are thus no density fluctuations in this gauge. However, one can make a transformation to a gauge in which $a_n = b_n = 0$. In this gauge the density fluctuation at the time that the wavelength comes within the horizon is

$$\langle(\Delta\rho/\rho)^2\rangle \approx y^2 \frac{\dot{\rho}_e^{\,2}}{\dot{a}_e^{\,2}\rho_e^{\,2}} \dot{a}_*^{\,2} . \tag{10.1}$$

Because y and $\dot{a}_*$ depend only logarithmically on the wavelength of the perturbations, this gives an almost scale-free spectrum of density fluctuations. These fluctuations can evolve according to the classical field equations to give rise to the formation of galaxies and all the other structure that we observe in the Universe. Thus all the complexities of the present state of the Universe have their origin in the ground-state fluctuations in the inhomogeneous modes and so arise from the Heisenberg uncertainty principle.

APPENDIX A: HARMONICS ON THE THREE-SPHERE

In this appendix we describe the properties of the scalar, vector, and tensor harmonics on the three-sphere S^3. The metric on S^3 is Ω_{ij} and so the line element is

$$dl^2 = \Omega_{ij}dx^i dx^j = d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2) . \tag{A1}$$

A vertical bar will denote covariant differentiation with respect to the metric Ω_{ij}. Indices i,j,k are raised and lowered using Ω_{ij}.

Scalar harmonics

The scalar spherical harmonics $Q^n_{lm}(\chi,\theta,\phi)$ are scalar eigenfunctions of the Laplacian operator on S^3. Thus, they satisfy the eigenvalue equation

$$Q^{(n)}{}_{|k}{}^{|k} = -(n^2-1)Q^{(n)}, \quad n = 1,2,3,\ldots . \tag{A2}$$

The most general solution to (A2), for given n, is a sum of solutions

$$Q^{(n)}(\chi,\theta,\phi) = \sum_{l=0}^{n-1}\sum_{m=-l}^{l} A^n_{lm} Q^n_{lm}(\chi,\theta,\phi) , \tag{A3}$$

where A^n_{lm} are a set of arbitrary constants. The Q^n_{lm} are given explicitly by

$$Q^n_{lm}(\chi,\theta,\phi) = \Pi^n_l(\chi) Y_{lm}(\theta,\phi) , \tag{A4}$$

where $Y_{lm}(\theta,\phi)$ are the usual harmonics on the two-sphere, S^2, and $\Pi^n_l(\chi)$ are the Fock harmonics.[19,20] The spherical harmonics Q^n_{lm} constitute a complete orthogonal set for the expansion of any scalar field on S^3.

Vector harmonics

The transverse vector harmonics $(S_i)^n_{lm}(\chi,\theta,\phi)$ are vector eigenfunctions of the Laplacian operator on S^3 which are transverse. That is, they satisfy the eigenvalue equation

$$S_i^{(n)}{}_{|k}{}^{|k} = -(n^2-2)S_i^{(n)}, \quad n = 2,3,4,\ldots \tag{A5}$$

and the transverse condition

$$S_i^{(n)|i} = 0 . \tag{A6}$$

The most general solution to (A5) and (A6) is a sum of solutions

$$S_i^{(n)}(\chi,\theta,\phi) = \sum_{l=1}^{n-1}\sum_{m=-l}^{l} B^n_{lm}(S_i)^n_{lm}(\chi,\theta,\phi) , \tag{A7}$$

where B^n_{lm} are a set of arbitrary constants. Explicit expressions for the $(S_i)^n_{lm}$ are given in Ref. 20 where it is also explained how they are classified as odd (o) or even (e) using a parity transformation. We thus have two linearly independent transverse vector harmonics S_i^o and S_i^e (n,l,m suppressed).

Using the scalar harmonics Q^n_{lm} we may construct a third vector harmonics $(P_i)^n_{lm}$. defined by (n,l,m suppressed)

$$P_i = \frac{1}{(n^2-1)} Q_{|i}, \quad n = 2,3,4,\ldots . \tag{A8}$$

It may be shown to satisfy

$$P_{i\,|k}{}^{|k} = -(n^2-3)P_i \quad \text{and} \quad P_i{}^{|i} = -Q . \tag{A9}$$

The three vector harmonics S_i^o, S_i^e, and P_i constitute a complete orthogonal set for the expansion of any vector field on S^3.

Tensor harmonics

The transverse traceless tensor harmonics $(G_{ij})^n_{lm}(\chi,\theta,\phi)$ are tensor eigenfunctions of the Laplacian operator on S^3 which are transverse and traceless. That is, they satisfy the eigenvalue equation

$$G_{ij}^{(n)}{}_{|k}{}^{|k} = -(n^2-3)G_{ij}^{(n)}, \quad n = 3,4,5,\ldots \tag{A10}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

and the transverse and traceless conditions

$$G_{ij}^{(n)}{}^{|i}=0,\quad G_i^{(n)i}=0\ . \tag{A11}$$

The most general solution to (A11) and (A12) is a sum of solutions

$$G_{ij}^{(n)}(\chi,\theta,\phi)=\sum_{l=2}^{n-1}\sum_{m=-l}^{l} C_{lm}^n (G_{ij})_{lm}^n(\chi,\theta,\phi)\ , \tag{A12}$$

where C_{lm}^n are a set of arbitrary constants. As in the vector case they may be classified as odd or even. Explicit expressions for $(G_{ij}^o)_{lm}^n$ and $(G_{ij}^e)_{lm}^n$ are given in Ref. 20.

Using the transverse vector harmonics $(S_i^o)_{lm}^n$ and $(S_i^e)_{lm}^n$, we may construct traceless tensor harmonics $(S_{ij}^o)_{lm}^n$ and $(S_{ij}^e)_{lm}^n$ defined, both for odd and even, by (n,l,m suppressed)

$$S_{ij}=S_{i\,|\,j}+S_{j\,|\,i} \tag{A13}$$

and thus $S_i{}^i=0$ since S_i is transverse. In addition, the S_{ij} may be shown to satisfy

$$S_{ij}{}^{|j}=-(n^2-4)S_i\ , \tag{A14}$$

$$S_{ij}{}^{|ij}=0\ , \tag{A15}$$

$$S_{ij\,|\,k}{}^{|k}=-(n^2-6)S_{ij}\ . \tag{A16}$$

Using the scalar harmonics Q_{lm}^n, we may construct two tensors $(Q_{ij})_{lm}^n$ and $(P_{ij})_{lm}^n$ defined by (n,l,m suppressed)

$$Q_{ij}=\tfrac{1}{3}\Omega_{ij}Q,\quad n=1,2,3 \tag{A17}$$

and

$$P_{ij}=\frac{1}{(n^2-1)}Q_{\,|\,ij}+\tfrac{1}{3}\Omega_{ij}Q,\quad n=2,3,4\ . \tag{A18}$$

The P_{ij} are traceless, $P_i{}^i=0$, and in addition, may be shown to satisfy

$$P_{ij}{}^{|j}=-\tfrac{2}{3}(n^2-4)P_i\ , \tag{A19}$$

$$P_{ij\,|\,k}{}^{|k}=-(n^2-7)P_{ij}\ , \tag{A20}$$

$$P_{ij}{}^{|ij}=\tfrac{2}{3}(n^2-4)Q\ . \tag{A21}$$

The six tensor harmonics Q_{ij}, P_{ij}, S_{ij}^o, S_{ij}^e, G_{ij}^o, and G_{ij}^e constitute a complete orthogonal set for the expansion of any symmetric second-rank tensor field on S^3.

Orthogonality and normalization

The normalization of the scalar, vector, and tensor harmonics is fixed by the orthogonality relations. We denote the integration measure on S^3 by $d\mu$. Thus

$$d\mu=d^3x(\det\Omega_{ij})^{1/2}=\sin^2\chi\sin\theta\,d\chi\,d\theta\,d\phi\ . \tag{A22}$$

The Q_{lm}^n are normalized so that

$$\int d\mu Q_{lm}^n Q_{l'm'}^{n'}=\delta^{nn'}\delta_{ll'}\delta_{mm'}\ . \tag{A23}$$

This implies

$$\int d\mu (P_i)_{lm}^n (P^i)_{l'm'}^{n'}=\frac{1}{(n^2-1)}\delta^{nn'}\delta_{ll'}\delta_{mm'} \tag{A24}$$

and

$$\int d\mu (P_{ij})_{lm}^n (P^{ij})_{l'm'}^{n'}=\frac{2(n^2-4)}{3(n^2-1)}\delta^{nn'}\delta_{ll'}\delta_{mm'}\ . \tag{A25}$$

The $(S_i)_{lm}^n$, both odd and even, are normalized so that

$$\int d\mu (S_i)_{lm}^n (S^i)_{l'm'}^{n'}=\delta^{nn'}\delta_{ll'}\delta_{mm'}\ . \tag{A26}$$

This implies

$$\int d\mu (S_{ij})_{lm}^n (S^{ij})_{l'm'}^{n'}=2(n^2-4)\delta^{nn'}\delta_{ll'}\delta_{mm'}\ . \tag{A27}$$

Finally, the $(G_{ij})_{lm}^n$, both odd and even, are normalized so that

$$\int d\mu (G_{ij})_{lm}^n (G^{ij})_{l'm'}^{n'}=\delta^{nn'}\delta_{ll'}\delta_{mm'}\ . \tag{A28}$$

The information given in this appendix about the spherical harmonics is all that is needed to perform the derivations presented in the main text. Further details may be found in Refs. 19 and 20.

APPENDIX B: ACTION AND FIELD EQUATIONS

The action (5.8) is

$$I=I_0(\alpha,\phi,N_0)+\sum_n I_n\ , \tag{B1}$$

where I_0 is the action of the unperturbed model (4.2):

$$I_0=-\tfrac{1}{2}\int dt\,N_0 e^{3\alpha}\left[\frac{\dot\alpha^2}{N_0{}^2}-e^{-2\alpha}-\frac{\dot\phi^2}{N_0{}^2}+m^2\phi^2\right]. \tag{B2}$$

I_n is quadratic in the perturbations and may be written

$$I_n=\int dt\,(L_g^n+L_m^n)\ , \tag{B3}$$

where

$$L_g^n=\tfrac{1}{2}e^\alpha N_0\left[\tfrac{1}{3}(n^2-\tfrac{5}{2})a_n{}^2+\frac{(n^2-7)}{3}\frac{(n^2-4)}{(n^2-1)}b_n{}^2-2(n^2-4)c_n{}^2-(n^2+1)d_n{}^2+\tfrac{2}{3}(n^2-4)a_nb_n\right.$$

$$\left.+g_n[\tfrac{2}{3}(n^2-4)b_n+\tfrac{2}{3}(n^2+\tfrac{1}{2})a_n]+\frac{1}{N_0{}^2}\left[-\frac{1}{3(n^2-1)}k_n{}^2+(n^2-4)j_n{}^2\right]\right]$$

$$+\frac{1}{2}\frac{e^{3\alpha}}{N_0}\left[-\dot a_n{}^2+\frac{(n^2-4)}{(n^2-1)}\dot b_n{}^2+(n^2-4)\dot c_n{}^2+\dot d_n{}^2\right.$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$+\dot{\alpha}\left[-2a_n\dot{a}_n+8\frac{(n^2-4)}{(n^2-1)}b_n\dot{b}_n+8(n^2-4)c_n\dot{c}_n+8d_n\dot{d}_n\right]$$

$$+\dot{\alpha}^2\left[-\tfrac{1}{2}a_n^2+6\frac{(n^2-4)}{(n^2-1)}b_n^2+6(n^2-4)c_n^2+6d_n^2\right]+g_n[2\dot{\alpha}\dot{a}_n+\dot{\alpha}^2(3a_n-g_n)]$$

$$+e^{-\alpha}\left[k_n\left[-\tfrac{2}{3}\dot{a}_n-\tfrac{2}{3}\frac{(n^2-4)}{(n^2-1)}\dot{b}_n+\tfrac{2}{3}\dot{\alpha}g_n\right]-2(n^2-4)\dot{c}_nj_n\right]\Bigg\} \tag{B4}$$

and

$$L_m^n=\tfrac{1}{2}N_0e^{3\alpha}\left[\frac{1}{N_0^2}(\dot{f}_n^2+6a_n\dot{f}_n\dot{\phi})-m^2(f_n^2+6a_nf_n\phi)-e^{-2\alpha}(n^2-1)f_n^2\right.$$

$$+\frac{3}{2}\left[\frac{\dot{\phi}^2}{N_0^2}-m^2\phi^2\right]\left[a_n^2-\frac{4(n^2-4)}{(n^2-1)}b_n^2-4(n^2-4)c_n^2-4d_n^2\right]+\frac{\dot{\phi}^2}{N_0^2}g_n^2$$

$$\left.-g_n\left[2m^2f_n\phi+3m^2a_n\phi^2+2\frac{\dot{f}_n\dot{\phi}}{N_0^2}+3\frac{a_n\dot{\phi}^2}{N_0^2}\right]-2\frac{e^{-\alpha}}{N_0^2}k_nf_n\dot{\phi}\right]. \tag{B5}$$

The full expressions for π_α and π_ϕ are

$$\pi_\alpha=\frac{e^{3\alpha}}{N_0}\left[-\dot{\alpha}+\sum_n\left[-a_n\dot{a}_n+\frac{4(n^2-4)}{(n^2-1)}b_n\dot{b}_n+4(n^2-4)c_n\dot{c}_n+4d_n\dot{d}_n\right]\right.$$

$$\left.+\dot{\alpha}\sum_n\left[-\tfrac{1}{2}a_n^2+\frac{6(n^2-4)}{(n^2-1)}b_n^2+6(n^2-4)c_n^2+6d_n^2\right]+\sum_n g_n[\dot{a}_n+\dot{\alpha}(3a_n-g_n)+\tfrac{1}{3}e^{-\alpha}k_n]\right], \tag{B6}$$

$$\pi_\phi=\frac{e^{3\alpha}}{N_0}\left\{\dot{\phi}+\sum_n\left[3a_n\dot{f}_n+\dot{\phi}\left[\tfrac{1}{2}a_n^2-\frac{4(n^2-4)}{(n^2-1)}b_n^2-4(n^2-4)c_n^2-4d_n^2\right]\right]\right.$$

$$\left.+\sum_n[\dot{\phi}g_n^2-g_n(\dot{f}_n+3a_n\dot{\phi})-e^{-\alpha}k_nf_n]\right\}. \tag{B7}$$

The classical field equations may be obtained from the action (B1) by varying with respect to each of the fields in turn. Variation with respect to α and ϕ gives two field equations, similar to those obtained in Sec. IV, but modified by terms quadratic in the perturbations:

$$N_0\frac{d}{dt}\left[\frac{1}{N_0}\frac{d\phi}{dt}\right]+3\frac{d\alpha}{dt}\frac{d\phi}{dt}+N_0^2m^2\phi=\text{quadratic terms}, \tag{B8}$$

$$N_0\frac{d}{dt}\left[\frac{\dot{\alpha}}{N_0}\right]+3\dot{\phi}^2-N_0^2e^{-2\alpha}-\tfrac{1}{2}(-\dot{\alpha}^2+\dot{\phi}^2-N_0^2e^{-2\alpha}+N_0^2m^2\phi^2)=\text{quadratic terms}. \tag{B9}$$

Variation with respect to the perturbations a_n, b_n, c_n, d_n, and f_n leads to five field equations:

$$N_0\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{a}_n}{N_0}\right]+\tfrac{1}{3}(n^2-4)N_0^2e^{\alpha}(a_n+b_n)+3e^{3\alpha}(\dot{\phi}\dot{f}_n-N_0^2m^2\phi f_n)=N_0^2[3e^{3\alpha}m^2\phi^2-\tfrac{1}{3}(n^2+2)e^{\alpha}]g_n$$

$$+e^{3\alpha}\dot{\alpha}\dot{g}_n-\tfrac{1}{3}N_0\frac{d}{dt}\left[e^{2\alpha}\frac{k_n}{N_0}\right], \tag{B10}$$

$$N_0\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{b}_n}{N_0}\right]-\tfrac{1}{3}(n^2-1)N_0^2e^{\alpha}(a_n+b_n)=\tfrac{1}{3}(n^2-1)N_0^2e^{\alpha}g_n+\tfrac{1}{3}N_0\frac{d}{dt}\left[e^{2\alpha}\frac{k_n}{N_0}\right], \tag{B11}$$

$$\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{c}_n}{N_0}\right]=\frac{d}{dt}\left[e^{2\alpha}\frac{j_n}{N_0}\right], \tag{B12}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$N_0\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{d}_n}{N_0}\right]+(n^2-1)N_0^2e^{\alpha}d_n=0\,, \tag{B13}$$

$$N_0\frac{d}{dt}\left[e^{3\alpha}\frac{\dot{f}_n}{N_0}\right]+3e^{3\alpha}\dot{\phi}\dot{a}_n+N_0^2[m^2e^{3\alpha}+(n^2-1)e^{\alpha}]f_n=e^{3\alpha}(-2N_0^2m^2\phi g_n+\dot{\phi}\dot{g}_n-e^{-\alpha}\dot{\phi}k_n)\,. \tag{B14}$$

In obtaining (B10)–(B14), the field equations (B8) and (B9) have been used and terms cubic in the perturbations have been dropped.

Variation with respect to the Lagrange multipliers k_n, j_n, g_n, and N_0 leads to a set of constraints. Variation with respect to k_n and j_n leads to the momentum constraints:

$$\dot{a}_n+\frac{(n^2-4)}{(n^2-1)}\dot{b}_n+3f_n\dot{\phi}=\dot{\alpha}g_n-\frac{e^{-\alpha}}{(n^2-1)}k_n\,, \tag{B15}$$

$$\dot{c}_n=e^{-\alpha}j_n\,. \tag{B16}$$

Variation with respect to g_n gives the linear Hamiltonian constraint:

$$3a_n(-\dot{\alpha}^2+\dot{\phi}^2)+2(\dot{\phi}\dot{f}_n-\dot{\alpha}\dot{a}_n)+N_0^2m^2(2f_n\phi+3a_n\phi^2)-\tfrac{2}{3}N_0^2e^{-2\alpha}[(n^2-4)b_n+(n^2+\tfrac{1}{2})a_n]$$
$$=\tfrac{2}{3}\dot{\alpha}e^{-\alpha}k_n+2g_n(-\dot{\alpha}^2+\dot{\phi}^2)\,. \tag{B17}$$

Finally, variation with respect to N_0 yields the Hamiltonian constraint, which we write as

$$\tfrac{1}{2}e^{3\alpha}\left[-\frac{\dot{\alpha}^2}{N_0^2}+\frac{\dot{\phi}^2}{N_0^2}-e^{-2\alpha}+m^2\phi^2\right]=\text{quadratic terms}\,. \tag{B18}$$

---

[1]A. H. Guth, Phys. Rev. D 23, 347 (1981).
[2]A. D. Linde, Phys. Lett. 108B, 389 (1982).
[3]S. W. Hawking and I. G. Moss, Phys. Lett. 110B, 35 (1982).
[4]A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. 48, 120 (1982).
[5]S. W. Hawking, Phys. Lett. 115B, 295 (1982).
[6]A. H. Guth and S. Y. Pi, Phys. Rev. Lett. 49, 1110 (1982).
[7]J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, Phys. Rev. D 28, 679 (1983).
[8]S. W. Hawking, Phys. Lett. B 150B, 339 (1985).
[9]V. A. Rubakov, M. V. Sazhin, and A. V. Veryaskin, Phys. Lett. 115B, 189 (1982).
[10]S. W. Hawking, Pontif. Accad. Sci. Varia 48, 563 (1982).
[11]J. B. Hartle and S. W. Hawking, Phys. Rev. D 28, 2960 (1983).
[12]S. W. Hawking, in *Relativity, Groups and Topology II*, Les Houches 1983, Session XL, edited by B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).
[13]S. W. Hawking, Nucl. Phys. B239, 257 (1984).
[14]S. W. Hawking and Z. C. Wu, Phys. Lett. 151B, 15 (1985).
[15]S. W. Hawking and D. N. Page, DAMTP report, 1984 (unpublished).
[16]S. W. Hawking and J. C. Luttrell, Nucl. Phys. B247, 250 (1984).
[17]L. P. Grishchuk, Zh. Eksp. Teor. Fiz. 67, 825 (1974) [Sov. Phys. JETP 40, 409 (1975)]; Ann. N.Y. Acad. Sci. 302, 439 (1977).
[18]A. A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. 30, 719 (1979) [JETP Lett. 30, 682 (1979)].
[19]E. M. Lifschitz and I. M. Khalatnikov, Adv. Phys. 12, 185 (1963).
[20]U. H. Gerlach and U. K. Sengupta, Phys. Rev. D 18, 1773 (1978).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PHYSICAL REVIEW D

PARTICLES AND FIELDS

THIRD SERIES, VOLUME 31, NUMBER 6 15 MARCH 1985

Quantum cosmology with a positive-definite action

Gary T. Horowitz

Department of Physics, University of California, Santa Barbara, California 93106

(Received 9 October 1984)

We argue that the indefiniteness of the Euclidean Einstein action is more serious in the cosmological context than in the asymptotically Euclidean context. To correct this, we consider a positive-definite action containing quadratic curvature terms. The physical states Ψ are now functions of both a three-metric q_{ab} and extrinsic curvature K_{ab}, and satisfy a differential equation analogous to the Wheeler-DeWitt equation. This equation has the form of a Schrödinger equation with q_{ab} playing the role of "time." By adopting Hartle and Hawking's boundary condition on the Euclidean function integral, we obtain a "preferred" solution to this equation. It is shown that in a simple minisuperspace model this wave function describes an inflationary universe.

I. INTRODUCTION

In a recent paper, Hartle and Hawking[1] have shown that there exists a "preferred" wave function in quantum gravity. This wave function has been further investigated by Hawking and Luttrell[2,3] who argue that it might reasonably represent the "quantum state of the universe." It is most conveniently defined in terms of the Euclidean functional-integral approach to quantum gravity.[4] In this approach, given a three-geometry $\hat{h}_{ab}$ one can formally define a wave function $\psi_{\hat{h}}[h_{ab}]$ by

$$\psi_{\hat{h}}[h_{ab}]=\int D[g_{ab}]e^{-I_{\rm GR}[g_{ab}]}, \tag{1.1}$$

where $I_{\rm GR}$ is the Euclidean Einstein action and the integral is over all Riemannian four-geometries which induce h_{ab} on one boundary and $\hat{h}_{ab}$ on the other. One can show that this wave function automatically satisfied the Wheeler-DeWitt equation, and hence represents a physical quantum state. Thus one obtains a quantum state for each choice of $\hat{h}_{ab}$. The preferred wave function mentioned above is obtained by letting $\hat{h}_{ab}$ degenerate to a point. To be more precise, if h_{ab} is a metric on a compact three-manifold Σ, then $\psi[h_{ab}]$ is defined by a Euclidean functional integral over all Riemannian four-metrics g_{ab} on manifolds M whose only boundary is Σ such that g_{ab} induces h_{ab} on Σ. If one thinks of $\psi_{\hat{h}}[h_{ab}]$ as the wave function "arising from $\hat{h}_{ab}$," then $\psi[h_{ab}]$ is the wave function "arising from nothing."[5] It is a possible candidate for the quantum state of the universe.

Unfortunately, there is a serious problem with this idea. The Euclidean form of the action for general relativity is not positive definite and hence the functional integral (1.1) is not well defined. To deal with this problem, it has been suggested that one must deform the contour of integration in the functional integral from an integral over real Euclidean metrics to an integral over complex metrics with $\mathrm{Re}\, I_{\rm GR}[g_{ab}]\geq 0$. Unfortunately, a completely satisfactory contour has not yet been found. Since this is the main motivation for what follows, we now review the situation in a little more detail.

The Euclidean Einstein action is[6]

$$I_{\rm GR}[g_{ab}]=\frac{1}{16\pi}\int_M(-R+2\Lambda)dV-\frac{1}{8\pi}\int_{\partial M}K\,d\Sigma\,, \tag{1.2}$$

where R is the scalar curvature of g_{ab}, Λ is a constant, and K is the trace of the extrinsic curvature of the boundary. The fact that $I_{\rm GR}$ is not positive definite has mainly been discussed in the context of asymptotically Euclidean metrics (and $\Lambda=0$). In this context, the following specific prescription was proposed[7] for dealing with this problem. Write a general asymptotically Euclidean metric $\tilde{g}_{ab}$ as $\tilde{g}_{ab}=\Omega^2 g_{ab}$ where Ω is a positive function and g_{ab} is a metric with zero scalar curvature. The positive-action theorem[8] states that asymptotically Euclidean metrics with zero scalar curvature have positive action. Thus, for $\Omega=1$, $I_{\rm GR}$ is positive definite. For $\Omega\neq 1$, the action includes a volume integral of $-(\nabla\Omega)^2$. To make this positive, one now writes the conformal factor as $\Omega=1+iy$ and does the functional integration over all functions y that vanish asymptotically.

There are at least two difficulties with this prescription. First, it is simply not true that a general asymptotically

 ©1985 The American Physical Society

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Euclidean metric can be conformally rescaled to one with zero scalar curvature. (A necessary condition is that the lowest eigenvalue of the conformally invariant Laplacian $-\nabla^2+\frac{1}{6}R$ should be positive.) Therefore this prescription completely omits a large class of possible configurations from the functional integral. A second difficulty concerns the presence of matter fields. Since the action for the matter fields involves the metric, it will no longer be positive when one rotates $\Omega=1+iy$. So the total action (matter plus gravity) can become negative.

The situation with regard to the indefiniteness of the Euclidean Einstein action is even worse in the cosmological context. To see this, fix a compact manifold M with boundary Σ and consider metrics $\tilde{g}_{ab}$ on M that induce a fixed metric h_{ab} on Σ. If we write $\tilde{g}_{ab}$ in the form

$$\tilde{g}_{ab}=\Omega^2 g_{ab} , \tag{1.3}$$

then from (1.2) we have

$$16\pi I_{\rm GR}(\tilde{g}_{ab})=\int_M[-\Omega^2R-6(\nabla\Omega)^2+2\Lambda\Omega^4]dV -2\int_{\partial M}\Omega^2K\,d\Sigma . \tag{1.4}$$

Even if we choose $\Lambda>0$, this action can clearly be negative. Let us try to follow the procedure used in the asymptotically Euclidean case. Since extrema of the action satisfy $R_{ab}=\Lambda g_{ab}$, it seems natural to require that $R=4\Lambda$. It is tempting now to simply rotate $\Omega\to i\Omega$ so that the volume term in the action becomes positive definite. However, this violates the boundary condition that the induced metric on Σ must be h_{ab}.

A second problem is that there is no analog of the positive-action theorem. For $\Omega=1$ and $R=4\Lambda$, the action becomes

$$8\pi I_{\rm GR}=-\Lambda V-\int K\,d\Sigma , \tag{1.5}$$

where V is the four-volume. Since the first term is negative definite and $K>0$ for spheres in a small neighborhood of a point this action can clearly be negative. In fact, instead of a positive-action theorem, one has the following:

Negative-action conjecture: Let M be a four-manifold with boundary Σ. Then for any metric on M with $R=4\Lambda$, the action $I_{\rm GR}$ defined by (1.5) is negative.

Intuitively, the idea is that in order for $I_{\rm GR}$ to be positive one needs $\int K<0$. But this means that one has to go past the "equator" of the geometry and by this time the volume is too big to be canceled. It seems unlikely that the action (1.5) is even bounded from below since one can construct "cylindrical" spaces with $R=4\Lambda$ and arbitrarily large volume.

In short, even if we ignore the fact that one cannot conformally rescale a general metric to one with $R=4\Lambda$,[9] and the fact that matter actions will no longer be positive, this procedure does not yield a positive-definite gravitational action.[10]

In this paper, the problem of an indefinite gravitational action is resolved by explicitly adding curvature-squared terms to the Einstein Lagrangian. The idea that the fundamental gravitational action is not the Einstein action but rather one involving quadratic terms in the curvature has received considerable attention recently for a variety of reasons.[11] Here, the main motivation is to improve the convergence of the Euclidean functional integral. We can then adopt exactly the same boundary condition that has been proposed for the Einstein Lagrangian to obtain a preferred "wave function of the universe."

In the past, theories described by curvature-squared Lagrangians were thought to suffer from instabilities and violation of unitarity. However, these conclusions were based entirely on a perturbation analysis off Minkowski spacetime. There is now growing evidence[11,12] that the perturbation theory is misleading, and that the full nonperturbative theory will be free of these difficulties.

The idea of computing the wave function of the universe in a theory described by a curvature-squared Lagrangian has also been discussed recently by Hawking and Luttrell.[3] However their approach is quite different from that adopted here. They add higher derivatives to mimic the effect of a massive scalar field in the gravitational Lagrangian. But their action is not positive definite.

What is the exact form of the action that we adopt? Perhaps the most attractive candidate for the action is simply to take the two independent curvature-squared Lagrangians

$$I_1=\int A(C_{abcd})^2+BR^2 , \tag{1.6}$$

where A and B are dimensionless coupling constants. This action was the starting point of the "induced gravity" program whose goal was to induce the scalar curvature term by quantum effects.[13] Unfortunately this program has recently run into difficulty because of the discovery of nonperturbative ambiguities in the value of the induced Newton's constant.[14] Therefore we must include the original Einstein action term:

$$I_2=\int A(C_{abcd})^2+BR^2-\frac{1}{16\pi}R . \tag{1.7}$$

This action, however, is *not* positive definite: Consider a metric whose curvature is small with R positive over a large volume. Then the curvature-squared terms will be negligible compared to the Einstein term which (for large enough volumes) can become arbitrarily negative. Thus we have to add a cosmological constant term Λ to make the action positive. So we are led to take the fundamental gravitational action to be

$$I=\tfrac{1}{4}\int A(C_{abcd})^2+B(R-4\Lambda)^2 , \tag{1.8}$$

where $32\pi\Lambda B=1$. Typically, one takes $B\simeq 1$ which implies that Λ is of order one in Planck units. This is, of course, much greater than the observational limits and is another manifestation of the well-known cosmological-constant problem.[15] If the Λ^2 term in (1.8) is somehow canceled in the effective action, then it is easy to show that any extrema of the resulting I whose typical radius of curvature is greater than the Planck length is an approximate solution to Einstein's equation. Thus one can recover general relativity in the "long wavelength" limit. It will be shown in the next section that one does *not* have to add a boundary term to this action.

The plan of this paper is the following. In Sec. II we begin by reviewing the canonical analysis of the Lorentzi-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

an theory described by the action I. This analysis shows that physical quantum states are described by wave functions ψ depending on a three-metric g_{ab} and extrinsic curvature K_{ab}, satisfying two differential equations coming from the classical constraints. The first equation implies that ψ is invariant under diffeomorphisms on the three-surface. The second has the form of a Schrödinger equation with q_{ab} playing the role of "time." We then consider the preferred wave function defined by the Euclidean functional integral. We will see that if one analytically continues the Euclidean extrinsic curvature to imaginary values (corresponding to real Lorentzian extrinsic curvature), then the wave function satisfies the constraint equations of the Lorentzian theory. In Sec. III this wave function is evaluated in a simple minisuperspace model. We find that ψ represents a superposition of classical de Sitter spacetimes. Hence it may be an appropriate description of an early inflationary phase in our universe.

II. COSMOLOGICAL WAVE FUNCTIONS

The wave function ψ is, of course, a function on configuration space. In order to identify the appropriate configuration space as well as to find the differential equations that ψ must satisfy, one must do a canonical analysis. We therefore begin this section with a discussion of the canonical theory. The second half of this section then examines the Euclidean functional integral which defines ψ and discusses its semiclassical evaluation.

A. Canonical analysis of the Lorentzian theory

The Lorentzian form of our fundamental gravitational action I is

$$\tilde{I}=-\tfrac{1}{4}\int AC_{abcd}C^{abcd}+B(R-4\Lambda)^2 . \tag{2.1}$$

This action has been cast into canonical form by Boulware.[16] We now briefly review his results.

Fix a compact three-manifold Σ. The canonical variables for the action $\tilde{I}$ can be taken to be the following: a three-metric q_{ab} on Σ and its conjugate momentum p^{ab} which is a symmetric second-rank tensor density that involves third time derivatives of the metric; and a tensor K_{ab} on Σ together with its conjugate momentum P^{ab} which is also a symmetric second-rank tensor density that involves second time derivatives of the metric. If we evolve $(q_{ab},K_{ab},P^{ab},p^{ab})$ to obtain a four-dimensional spacetime, then all these fields except p^{ab} acquire a simple geometric interpretation: q_{ab} is the induced metric on the surface Σ, K_{ab} is its extrinsic curvature, and P^{ab} is simply related to the four-dimensional curvature by

$$P^{ab}=-2q^{1/2}(2AE^{ab}+Bq^{ab}R) , \tag{2.2}$$

where $q=\det q_{ab}$, $E^{ab}=C^{ambn}t_m t_n$ is the electric part of the Weyl tensor with respect to the unit normal to Σ, and R is the four-dimensional scalar curvature. (Boulware in fact chooses $-P^{ab}$ as a "position variable" and K_{ab} as the conjugate momentum. The resulting formalism is of course completely equivalent. However, we will see in Sec. II B that, from the standpoint of quantization, the q_{ab}, K_{ab} representation is more convenient.)

The variables $(q_{ab},K_{ab},P^{ab},p^{ab})$ cannot be chosen independently. Just as in general relativity there are constraints arising from the invariance of the action under diffeomorphisms. These constraints are

$$\begin{aligned}C=&K_{ab}p^{ab}-\frac{1}{8Aq^{1/2}}(P^T_{ab})^2-\frac{P^2}{144Bq^{1/2}}\\&+\tfrac{1}{2}P^{ab}(K_{ab}K+{}^3R_{ab})-\tfrac{1}{4}P({}^3R-K^{ab}K_{ab}+K^2)\\&+\tfrac{1}{2}D_aD_bP^{ab}-Aq^{1/2}(C_{abcd}t^d)^2\\&-2B\Lambda q^{1/2}({}^3R+K_{ab}K^{ab}-K^2)+4\Lambda^2Bq^{1/2}=0\end{aligned} \tag{2.3}$$

and

$$C_m=-\tfrac{1}{2}P^{ab}D_mK_{ab}+D_a(K_{bm}P^{ab})+D_ap^a_m=0 , \tag{2.4}$$

where $(\ )^2$ means square the tensor with the metric q^{ab} [so $(C_{abcd}t^d)^2$ involves just spatial derivatives of K_{ab}] and P^T_{ab}, and P are the tracefree and trace parts of P_{ab}.

In the standard canonical approach to quantization the physical states of the system are represented by functions on configuration space $\psi(q_{ab},K_{ab})$ satisfying the operator version of the constraints:

$$\mathbf{C}\psi=0 , \tag{2.5}$$

$$\mathbf{C}_m\psi=0 , \tag{2.6}$$

where $\mathbf{C}$ and $\mathbf{C}_m$ are obtained by replacing p^{ab} by $(1/i)\delta/\delta q_{ab}$ and P^{ab} by $(1/i)\delta/\delta K_{ab}$ in (2.3) and (2.4). These constraint operators have similar factor-ordering and regularization problems as the analogous operators arising from the Einstein action.

The vector constraint (2.4) has exactly the same interpretation as the analogous constraint in general relativity. Classically, it generates diffeomorphisms on the three-surface. Quantum mechanically, it requires that the wave function depend only on the equivalence classes $\{(q_{ab},K_{ab})\}$ where two pairs (q_{ab},K_{ab}) and (q'_{ab},K'_{ab}) are said to be equivalent if they are related by a diffeomorphism on the three-surface Σ. This is easily seen as follows. Let N^m be any vector field on Σ. Then from (2.4) and a simple integration by parts we have

$$\begin{aligned}\int_\Sigma N^mC_m=&\int_\Sigma[-\tfrac{1}{2}P^{ab}N^mD_mK_{ab}\\&-P^{ab}K_{bm}D_aN^m-p^{ab}D_aN_b]\\=&-\tfrac{1}{2}\int_\Sigma[(\mathcal{L}_NK_{ab})P^{ab}+(\mathcal{L}_Nq_{ab})p^{ab}] .\end{aligned} \tag{2.7}$$

Therefore, replacing P^{ab} by $(1/i)\delta/\delta K_{ab}$ and P^{ab} by $(1/i)\delta/\delta q_{ab}$ we see that $\int_\Sigma N^mC_m\psi=0$ says that ψ is unchanged if we simultaneously change q_{ab} and K_{ab} by the diffeomorphism generated by N^m.

The scaler constraint (2.5) is the analog of the Wheeler-DeWitt equation for the Einstein action. Interestingly enough, the form of this equation is quite different. The Wheeler-DeWitt equation resembles a wave equation on the configuration space. Equation (2.5) has the general structure of a Schrödinger equation with the three-metric q_{ab} playing the role of time. With the sim-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

plest choice of factor ordering, this equation takes the following form:

$$K_{ab}\, i\frac{\delta}{\delta q_{ab}} = G_{abcd}\frac{\delta}{\delta K_{ab}}\frac{\delta}{\delta K_{cd}} + A_{ab}\frac{\delta}{\delta K_{ab}} + V\,, \quad (2.8)$$

where G_{abcd} is a "metric" which depends only on q_{ab} and is positive definite. Although q_{ab} probably contains information about physical fields as well as the time, the fact that it naturally appears in a first-order form might have a significant effect in improving our understanding of time in quantum gravity. Notice that since q_{ab} and K_{ab} are independent variables the quadratic term is independent of the usual factor-ordering problems present in the analogous term in the Wheeler-DeWitt equation. However the linear-momentum term does have factor-ordering ambiguities since A_{ab} as well as V depend on both q_{ab} and K_{ab}. The precise expressions for G_{abcd}, A_{ab} and V can be read off from (2.3).[17]

The fact that wave functions in this theory depend on K_{ab} as well as q_{ab} has an important consequence. Penrose has argued[18] from considerations of entropy that the "big bang" should be very different from the "big crunch." In a theory where the quantum states are functions only of the three-metric, it is difficult to understand how this asymmetry could come about. However in this theory, the structure of the wave function for small three-volume and $K>0$ can be quite different from its structure for small three-volume and $K<0$. If the preferred wave function defined by the Euclidean functional integral has this property, then this might provide a possible "explanation" for the observed time asymmetry of our universe.

B. Euclidean functional integral

The canonical analysis discussed above shows that the appropriate configuration space for the theory described by the action (2.1) consists of pairs (q_{ab},K_{ab}) where q_{ab} is a metric and K_{ab} a symmetric tensor on a compact three-manifold Σ. We can now obtain a preferred Euclidean wave function ψ_E by adopting the same boundary condition on the functional integral as used for the Einstein Lagrangian. That is

$$\psi_E(q_{ab},K_{ab}) = \int D[g_{ab}]e^{-I[g]}\,, \quad (2.9)$$

where the integral is over all Euclidean metrics g_{ab} on manifolds M whose boundary is Σ such that the induced metric and extrinsic curvature on Σ are q_{ab} and K_{ab}, respectively.

Before discussing the relation between ψ_E and the physical states of the theory, i.e., wave functions satisfying (2.5) and (2.6), we first consider the question of whether the action I given by (1.8) should be supplemented by boundary terms. For the Einstein action there are two separate arguments which require that a boundary term be added to the integral of the scalar curvature. In the first argument, the boundary term is obtained by demanding that solutions of the classical field equation be extrema of the action under all perturbations that vanish on the boundary, i.e., $\delta q_{ab}=0$. In order to satisfy this condition one has to "cancel" the second derivative terms in R by adding the integral of the extrinsic curvature. The analogous condition for the curvature-squared action I is that solutions be extrema of I under all perturbations satisfying $\delta q_{ab}=\delta K_{ab}=0$ on the boundary. This is because one expects to get a unique classical solution only if one fixes both q_{ab} and K_{ab} on the boundary. But, it is easy to verify that solutions of the classical field equation are indeed extrema of I (with no extra boundary terms added) under all perturbations satisfying these conditions.

The second argument, which leads to the same boundary term for the Einstein action, is the following. Consider a transition from an initial configuration (q_1,K_1) on a surface Σ_1 to a configuration (q_2,K_2) on a surface Σ_2, followed by a transition from (q_2,K_2) to a final configuration (q_3,K_3) on a surface Σ_3. One expects that the amplitude to go from the initial to the final configuration should be obtained by integrating over all configurations on the intermediate surface Σ_2. This requires

$$I[g_1+g_2]=I[g_1]+I[g_2]\,, \quad (2.10)$$

where g_1 is a metric that induces (q_1,K_1) on Σ_1 and (q_2,K_2) on Σ_2, g_2 is a metric that induces (q_2,K_2) on Σ_2 and (q_3,K_3) on Σ_3, and g_1+g_2 is the metric obtained by taking the union of the two. Since q_1 and q_2 induce the same metric *and intrinsic curvature* on Σ_2 the four-dimensional curvature of g_1+g_2 will not contain any δ functions (although it may be discontinuous). Hence *condition (2.10) is automatically satisfied by the action (1.8) without having to add any boundary terms.* Since the configuration space for the Einstein Lagrangian is smaller, the three metrics induced on Σ_2 by g_1 and g_2 will still agree; but the extrinsic curvatures need not. This results in a δ function in the Ricci curvature proportional to the difference of the extrinsic curvatures on Σ_2. Hence condition (2.10) will hold for I_{GR} only if one includes the surface term.

To summarize, in both cases the action I given by (1.8) satisfies the desired properties without having to add any surface terms.[19]

We now turn to the relation between ψ_E and the physical states of the theory, i.e., wave functions annihilated by the constant operators. Since ψ_E is a function of Euclidean extrinsic curvature, one would not expect it to satisfy (2.5) and (2.6). If we choose coordinates near Σ so that the Lorentzian extrinsic curvature is $K^L_{ab}=\partial q_{ab}/\partial t$, then the Euclidean action can be obtained from the Lorentzian action by formally setting $\tau=it$. The Euclidean extrinsic curvature is then

$$K^E_{ab}=\frac{\partial q_{ab}}{\partial \tau}=-i\frac{\partial q_{ab}}{\partial t}=-iK^L_{ab}\,. \quad (2.11)$$

Thus we expect that

$$\psi(q_{ab},K^L_{ab})\equiv\psi_E(q_{ab},-iK^L_{ab}) \quad (2.12)$$

will satisfy the Lorentzian constraint equations. We now show that, at least formally, this is indeed the case.

Let M be a manifold containing a compact three-dimensional submanifold Σ *(not* on its boundary). Given q_{ab} and K^E_{ab} on Σ we define $\psi_E(q_{ab},K^E_{ab})$ by the Euclidean functional integral (2.9). Since there is no external measure of time, ψ_E does not depend on where Σ is placed in M. If we took a different submanifold Σ' and put the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

same fields q_{ab}, K^E_{ab} on Σ', then the value of ψ_E is left unchanged. But by the "same fields" we mean that there exists a diffeomorphism $\phi:\Sigma\to\Sigma'$ which takes the fields on Σ to the fields on Σ'. Hence ψ_E must be invariant under diffeomorphisms. (One can think of a diffeomorphism as just corresponding to a "change of variables" in the functional integral.) Now consider an infinitesimal diffeomorphism generated by a vector field ξ^a. Taking the variation of both sides of (2.9) and assuming the measure is invariant we get

$$0=\int D[g]\left[\int_M C^{ab}\nabla_{(a}\xi_{b)}\right]e^{-I[g]}\,. \tag{2.13}$$

where $C^{ab}\equiv\delta I/\delta g_{ab}$ is the classical (Euclidean) field equation for the action I. Using the fact that $\nabla_a C^{ab}=0$, we can integrate by parts inside the large parentheses to obtain

$$0=\int D[g]\left[\int_\Sigma C^{ab}\xi_a d\Sigma_b\right]e^{-I[g]}\,. \tag{2.14}$$

The quantity in large parentheses is just the (Euclidean) constraint equations. Equation (2.14) shows that the operator versions of these constraints automatically annihilate the wave function.

Unlike the case of the Einstein Lagrangian, the operator version of the Euclidean constraints is *not* the same as the operator version of the Lorentzian constraints. This is easily seen as follows. The Euclidean constraints are obtained from the Lorentzian constraints (2.3) and (2.4) by setting $K^L_{ab}=iK^E_{ab}$, $p^{ab}_L=ip^{ab}_E$, and leaving q_{ab} and P^{ab} unchanged. The Euclidean constraint operators are obtained from the constraints by replacing $P^{ab}\to-\delta/\delta K^E_{ab}$ and $p^{ab}_E\to-\delta/\delta q_{ab}$. The net effect is thus to send $q_{ab}\to q_{ab}$ $P^{ab}_L\to(1/i)\delta/\delta q_{ab}$ (which agrees with the Lorentzian operator equation) and $K^L_{ab}\to iK^E_{ab}$, $P^{ab}_L\to-\delta/\delta K^E_{ab}$ (which does not). Thus to obtain a solution of the Lorentzian operator constraints we must analytically continue the Euclidean wave function to imaginary values of the extrinsic curvature:

$$\psi[q_{ab},K^L_{ab}]=\psi_E[q_{ab},-iK^L_{ab}]\,. \tag{2.15}$$

This procedure will yield a solution to the constraints (2.5) and (2.6) for any wave function defined by the Euclidean functional integral. By adopting Hartle and Hawking's preferred boundary condition,[1] we obtain a unique physical state which might represent the quantum state of our universe.

One cannot, of course, calculate the functional integral (2.9) exactly. (In fact, the precise form of the measure is still an open question.) However, one can get some information about ψ from a semiclassical approximation. Since we are interested in ψ_E only for imaginary values of the extrinsic curvature, the semiclassical approximation is somewhat complicated. For example, it is not sufficient to minimize the Euclidean action with q_{ab} and K^E_{ab} fixed on the boundary and then analytically continue $\psi_E(q_{ab},K^E_{ab})=e^{-I_{\min}[g_{ab}]}$. This is because the analytic continuation may cross "Stokes lines" where the asymptotic form of the integral (for small $\hbar$) changes character. Instead, one must do a steepest-descents analysis in which one looks for complex extrema of the action with q_{ab} and $-iK^L_{ab}$ fixed on the boundary. Then, if one can deform the original contour of integration into a steepest-descents contour passing through this extrema, the semiclassical approximation to ψ is

$$\psi(q_{ab},K^L_{ab})\approx N\,e^{-I[g_{ab}]}\,, \tag{2.16}$$

where g_{ab} is the complex extremum and N includes contributions from quadratic fluctuations.

We now briefly explain what is meant by complex extrema of the action. Given a real manifold, one can always complexify the tangent space and introduce complex tensor fields. (Since we want to consider the boundary of our manifold $\partial M=\Sigma$ to be a three-dimensional submanifold, it is better not to view M as a complex two-manifold, even when this is possible.) A complex metric is simply an invertible, symmetric second-rank complex tensor field. There exists a unique, torsion-free covariant derivative compatible with each complex metric, and one can write down the field equations just as for real metrics. The only subtlety comes at the boundary. The unit normal to Σ can be defined by taking any function f such that $f=0$, $\nabla_a f\neq0$ on Σ and setting

$$n_a=\frac{\nabla_a f}{(g^{mn}\nabla_m f\nabla_n f)^{1/2}}\,. \tag{2.17}$$

The problem is that the square root in the denominator is only defined up to an overall sign. In other words, one cannot distinguish the "outgoing" and "ingoing" normals for a general complex metric. This does not affect the definition of q_{ab} but it does affect the definition of K_{ab}. In our case this ambiguity is resolved as follows. Since we begin with an integral over real Euclidean metrics, there is a well-defined outgoing normal. We now require that as we deform the contour of integration, the unit normal changes continuously. This gives a well-defined definition of K_{ab}.

We conclude this section with a few remarks about another representation for the quantum theory based on the action (2.1). Instead of choosing (q_{ab},K_{ab}) to be "position variables," we could have chosen the configuration space to consist of pairs (q_{ab},P^{ab}). Although the resulting theory should be completely equivalent, the (q_{ab},P^{ab}) representation is less desirable for at least three reasons.

First, the scalar constraint equation (2.5) changes its character completely: Since the only quadratic momentum term in the (q_{ab},P^{ab}) representation is $K_{ab}p^{ab}$, the analog of the Wheeler-DeWitt equation would no longer resemble Schrödinger's equation but rather a second-order equation in a space whose metric has six positive and six negative eigenvalues at each point. Second, the condition for the composition of amplitudes in the (q_{ab},P^{ab}) representation is more complicated than $I[g_1+g_2]=I[g_1]+I[g_2]$. In fact, this latter condition can no longer be satisfied for *any* choice of boundary terms. This is because the extrinsic curvature need not be continuous across the boundary, so the four-dimensional curvature computed from g_1+g_2 may contain δ-function contributions. Since the action involves the square of the curvature, the left-hand side can diverge when the right-hand side is finite. Finally, it is difficult to obtain nontrivial information

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

about $\psi(q_{ab},P^{ab})$ from a semiclassical approximation. One can show that the wave function defined by the Euclidean functional integral with q_{ab} and P^{ab} held fixed on the boundary automatically satisfies the Lorentzian constraint without any analytic continuation. Although this would normally be a significant advantage, in our case it is not, since the metric which minimizes the action holding q_{ab} and P^{ab} fixed on the boundary is in fact independent of P^{ab}. The curvature becomes discontinuous on the boundary.

A simple example will illustrate this last point. Let φ be a function on the unit ball B in E^4 and set

$$S[\varphi]=\int_B(\nabla^2\varphi)^2 . \tag{2.18}$$

We wish to minimize S subject to the condition that on the boundary of B, $\varphi=0$ and $\nabla^2\varphi=K$ where K is some constant. Consider a sequence of functions φ_{r_0} which are constant for $r<r_0<1$ and then "bend" to satisfy $\nabla^2\varphi_{r_0}=K$ at $r=1$. One possibility is

$$\varphi_{r_0}(r)=\begin{cases}-K(1-r_0)^2/(8-6r_0), & 0\le r\le r_0 ,\\ K(r-1)(r+1-2r_0)/(8-6r_0), & r_0\le r\le 1 .\end{cases} \tag{2.19}$$

It is easy to see that both φ_{r_0} and φ'_{r_0} are continuous at $r=r_0$ and φ_{r_0} satisfies the boundary conditions for all r_0. The action $S[\varphi_{r_0}]$ is clearly minimized by taking the limit $r_0\to 1$. In this limit, the action approaches zero. In particular it is independent of the value of K. Since the action I does not contain any terms involving derivatives of the curvature, a similar thing can happen here. Thus, nontrivial information about $\psi(q_{ab},P^{ab})$ can only be obtained by computing the quadratic fluctuations.

III. MINISUPERSPACE MODEL

In this section we apply the general formalism discussed in Sec. II to a simple minisuperspace model. The model is obtained by restricting consideration only to metrics with isotropic three-surfaces. Since metrics of this type are all conformally flat, the Weyl squared term in the action I (1.8) will play no role. Furthermore, we set $\Lambda=0$. Although a physically realistic theory probably needs $\Lambda>0$, to include a positive Newton's constant, the choice $\Lambda=0$ is made for the following reason. It simplifies the semiclassical evaluation of the wave function considerably, and yields some information about the $\Lambda>0$ case. This is because Λ is negligible for certain regions of the configuration space. A complete analysis of the case $\Lambda>0$ will appear in a forthcoming paper.[20] For now we consider the simpler theory, and show that with $\Lambda=0$, the resulting wave function describes an inflationary universe.

We begin this section with a canonical analysis of the Lorentzian theory. Then we evaluate the functional integral semiclassically and interpret the resulting wave function.

A. Canonical analysis

We write the spacetime metric in the form

$$ds^2=e^{2\alpha(\eta)}(-d\eta^2+d\Omega_3{}^2) , \tag{3.1}$$

where $d\Omega_3{}^2$ is the metric on a unit three-sphere. The four-dimensional scalar curvature is then given by

$$R=6e^{-2\alpha}(1+\alpha'^2+\alpha'') , \tag{3.2}$$

where a prime denotes $d/d\eta$. The Lorentzian action is (2.1)

$$\tilde{I}=\frac{-B}{4}\int R^2dV$$

$$=-18B\pi^2\int(1+\alpha'^2+\alpha'')^2d\eta . \tag{3.3}$$

To cast this fourth-order action into canonical form we must first vary $\tilde{I}$ with respect to the highest derivatives and set the results equal to a new variable Q:

$$Q\equiv-\frac{\delta\tilde{I}}{\delta\alpha''}=36B\pi^2(1+\alpha'^2+\alpha'') . \tag{3.4}$$

Solving for α'' in terms of Q we next write the action in the form

$$\tilde{I}=\int(\alpha'Q'-\bar{H})d\eta , \tag{3.5}$$

where

$$\bar{H}(\alpha,\alpha',Q)=(1+\alpha'^2)Q-Q^2/m \tag{3.6}$$

and we have set $m\equiv(72\pi^2B)$. The momenta conjugate to α and Q are now defined in the usual manner:

$$P_Q=\frac{\delta\tilde{I}}{\delta Q'}=\alpha' , \tag{3.7a}$$

$$P_\alpha=\frac{\delta\tilde{I}}{\delta\alpha'}=Q'-2\alpha'Q . \tag{3.7b}$$

The Hamiltonian is then

$$H=P_\alpha Q'+P_\alpha\alpha'-L$$

$$=P_\alpha P_Q+(1+P_Q{}^2)Q-Q^2/m . \tag{3.8}$$

Classically, this Hamiltonian is constrained to vanish. In the quantum theory we require that the Hamiltonian operator must annihilate the wave function. As discussed in the last section we want to quantize in a representation where α and $P_Q=\alpha'$ are the configuration variables.[21] Therefore, we replace P_Q with x and Q with $-P_x$ to obtain

$$H=xP_\alpha-(1+x^2)P_x-P_x{}^2/m . \tag{3.9}$$

To obtain the Hamiltonian operator we replace P_α with $(1/i)\partial/\partial\alpha$ and P_x with $(1/i)\partial/\partial x$. Thus physical states of the theory must satisfy

$$xi\frac{\partial\psi}{\partial t}=\frac{1\partial^2\psi}{m\partial x^2}+i\frac{\partial\psi}{\partial x}+ix^{2-p}\frac{\partial(x^p\psi)}{\partial x} , \tag{3.10}$$

where we have replaced α with the more suggestive variable t and written the last term with an arbitrary parameter p reflecting ambiguities in factor ordering. *This equation is just the ordinary Schrödinger equation for a one-dimensional system governed by the time-independent Hamiltonian*

$$H_0(x,p)=\frac{-1}{x}\left[\frac{p^2}{m}+(1+x^2)p\right] . \tag{3.11}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

The relation between H_0 and H is the following. Since H is independent of α, P_α is constant. H_0 is simply $-P_\alpha$. The equations of motion generated by H_0 are completely equivalent to those generated by H. The only difference is that for H_0, the solutions are parametrized by t whereas for H, they are parametrized by η. Note that H_0 is not bounded from below.

Given any function $\psi(t_0,x)$ one can in principle evolve (3.10) to obtain an allowed quantum state of the system. In the next section we find which quantum state corresponds to placing Hartle and Hawking's boundary condition on the Euclidean functional integral.

B. Semiclassical evaluation of functional integral

We now consider positive-definite metrics of the form:

$$ds^2=e^{2\alpha(\eta)}(d\eta^2+d\Omega_3^2)\,. \tag{3.12}$$

The scalar curvature is

$$R=6\,e^{-2\alpha}(1-\alpha'^2-\alpha'') \tag{3.13}$$

[cf. (3.2)]. Recall that to obtain the "wave function of the universe" one must first compute

$$\psi_E(\alpha_0,\alpha_0')=\int D[\alpha]e^{-I[\alpha]}\,, \tag{3.14}$$

where the Euclidean action is

$$I[\alpha]=\frac{B}{4}\int R^2dV \tag{3.15}$$

and the integral is over all metrics which are regular at the origin and induce α_0 and α_0' on the boundary. To see what regularity at the origin corresponds to for (3.12), we briefly consider an alternative form of the metric

$$ds^2=a^2(r)(dr^2+r^2d\Omega_3^2)\,. \tag{3.16}$$

This metric is clearly regular at the origin if $a(r)=a(0)+O(r^2)$. Now comparing (3.12) and (3.16) we obtain

$$\frac{dr}{r}=d\eta,\quad 1+\frac{r}{a}\frac{da}{dr}=\alpha'\,. \tag{3.17}$$

Therefore the origin corresponds to $\eta=-\infty$ and regularity at the origin requires that for $\eta\ll 0$, $\alpha'\approx 1+O(e^{2\eta})$. After computing $\psi_E(\alpha_0,\alpha_0')$ one must then analytically continue α_0' to $-ix_0$ to obtain a solution to (3.10).

As we discussed earlier, in the semiclassical approximation, it is not sufficient to minimize the Euclidean action, and then analytically continue $\psi_E(\alpha_0,\alpha_0')=e^{-I_{\min}(\alpha_0,\alpha_0')}$ from real α_0' to imaginary values. Instead, one must do a steepest-descents analysis. This involves finding complex extrema of the action which satisfy the boundary conditions (denoting the boundary by $\eta=0$):

$$\begin{aligned}&\alpha'(\eta)=1+O(e^{2\eta})\ \text{ for } \eta\ll 0\,,\\ &\alpha(0)=\alpha_0\,,\\ &\alpha'(0)=-ix_0\,.\end{aligned} \tag{3.18}$$

If the original contour of integration can be deformed into a steepest-descents contour passing through the extrema $\alpha(\eta)$, then the semiclassical approximation to ψ is

$$\psi(\alpha_0,x_0)\approx N\,e^{-I[\alpha]}\,, \tag{3.19}$$

where N includes contributions from quadratic fluctuations. In the following we will concentrate on just the exponent. We now find the complex extrema of the action I satisfying the boundary conditions (3.18).

Setting the variation of I with respect to α equal to zero we obtain the field equation

$$[(1-\alpha'^2-\alpha'')'-2\alpha'(1-\alpha'^2-\alpha'')]'=0\,. \tag{3.20}$$

Using (3.13) this simplifies to

$$e^{2\alpha}R'=\text{constant}\,. \tag{3.21}$$

Since the scalar curvature must be finite at the origin where $\alpha\to-\infty$, this implies that $R=$constant. Hence, *all (isotropic) extrema of I have constant scalar curvature.* This result clearly holds for complex α as well as real α.

Since we have reduced the fourth-order equation (3.20) to the second-order equation $R=$constant, it is not clear that there will exist solutions satisfying our boundary conditions. Indeed, in the case of real α, one can show that all metrics with a regular origin and $R=$constant are in fact spaces of constant curvature. Since $\alpha'>-1$ for all spaces of constant curvature, there do not exist solutions with $\alpha_0'<-1$ on the boundary. (If one tries to minimize the action with $\alpha_0'<-1$ then the volume becomes unbounded, i.e., "the solution goes off to infinity.") However, we will now show that there *do* exist complex solutions satisfying (3.18) for all α_0,x_0.

We start with the metric for a real four-sphere:

$$e^{-\alpha}=M\cosh(\eta+c)\,, \tag{3.22}$$

where $M>0$ and c are constants. The scalar curvature is related to M by

$$R=12M^2\,. \tag{3.23}$$

Taking the derivative of both sides of (3.22) we obtain

$$\alpha'=-\tanh(\eta+c) \tag{3.24}$$

which shows that the boundary condition as $\eta\to-\infty$ is satisfied for all M and c. For real c, $|\alpha'|<1$ for all η. However, we now let $c=i\theta$ where $-\pi/2<\theta<\pi/2$. Then

$$e^{-\alpha}=M\cosh(\eta+i\theta) \tag{3.25}$$

has $R=$constant and satisfies the boundary conditions (3.18) provided

$$e^{-\alpha_0}=M\cos\theta\,,\quad x_0=\tan\theta\,. \tag{3.26}$$

We thus obtain complex extrema of I satisfying (3.18) for all α_0, x_0.

We could have obtained exactly the same solutions by starting with a real space of constant negative curvature, as we now show. The metric for a space of constant negative curvature is

$$e^{-\alpha}=-N\sinh(\eta-d)\,, \tag{3.27}$$

where $N>0$ and d are constants. (The point $\eta=d$ corresponds to infinity.) The scalar curvature is now $R=-12N^2$ and

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$\alpha' = -\coth(\eta - d) . \tag{3.28}$$

Setting $d = i\varphi$ $(0 < \varphi < \pi)$ we obtain a solution satisfying the boundary conditions (3.18) provided

$$e^{-\alpha_0} = iN \sin\varphi, \quad x_0 = \cot\varphi . \tag{3.29}$$

But the first condition says that N must be imaginary. If we set $N = -iM$ and $\varphi = \theta + \pi/2$, then this solution becomes identical to the one given by (3.25) and (3.26).

We now compute the action for these solutions:

$$I = 18\pi^2 B \int_{-\infty}^{0} (1 - \alpha'^2 - \alpha'')^2 d\eta . \tag{3.30}$$

From (3.24) we have

$$\alpha'' = (\alpha'^2 - 1) . \tag{3.31}$$

So

$$\begin{aligned} I &= 72\pi^2 B \int_{-\infty}^{0} (\alpha'^2 - 1)^2 d\eta \\ &= m \int_{1}^{-ix_0} (\alpha'^2 - 1) d\alpha' \\ &= m[i(x_0 + x_0^3/3) + \tfrac{2}{3}] , \end{aligned} \tag{3.32}$$

where we have again set $m = 72\pi^2 B$.

Now fix α_0 and x_0. It seems likely that the solution given by (3.25) and (3.26) is the only exact solution satisfying the prescribed boundary conditions. Although a detailed steepest-descents analysis is difficult for the infinite-dimensional path integral (3.14), it is reasonable to expect that the original contour of integration can be deformed to pass over this extrema. Thus, in the semiclassical approximation,

$$\psi(\alpha_0, x_0) \approx e^{-I[\alpha]} , \tag{3.33}$$

where I is given by (3.32). Replacing α_0 with t and x_0 with x we obtain our final result:

$$\psi(x,t) = e^{-im(x + x^3/3)} , \tag{3.34}$$

where we have dropped the irrelevant constant $2m/3$ from the action. Notice that the wave function is time independent. This is a direct result of the fact that the action is scale invariant, and hence is unchanged if a constant is added to α.

We now ask if ψ satisfies the Hamiltonian constraint of the canonical theory (3.10). Notice that

$$\frac{1}{i}\frac{\partial \psi}{\partial x} = -m(1 + x^2)\psi . \tag{3.35}$$

If we choose the factor-ordering parameter $p = 2$ in (3.10) then the constraint equation becomes

$$-xi\frac{\partial \psi}{\partial t} = \frac{1}{i}\frac{\partial}{\partial x}\left[\frac{1}{im}\frac{\partial}{\partial x} + (1 + x^2)\right]\psi . \tag{3.36}$$

Therefore ψ is an exact solution of the canonical constraint equation. It is perhaps surprising that the semiclassical approximation to the wave function satisfies the exact constraint equation. This is probably a result of the high degree of symmetry imposed on the metric together with scale invariance of the action. The fact that a certain choice of factor ordering is picked out is *not* surprising. It is (presumably) just the factor ordering which corresponds to choosing a scale-invariant measure in the path integral.[22] Notice that even though the one-dimensional Hamiltonian H_0 is unbounded from below, the wave function obtained by imposing Hartle and Hawking's boundary condition on the path integral corresponds to a zero-energy eigenstate.

The wave function ψ has a simple physical interpretation. It satisfies the operator equivalent of [see (3.35)]

$$p = -m(1 + x^2) , \tag{3.37}$$

Recall from (3.4) that $p = -Q = -mRe^{2\alpha}/12$ and $x = \alpha'$. Therefore, classically, (3.37) becomes

$$R = 12\, e^{-2\alpha}(1 + \alpha'^2) . \tag{3.38}$$

Taking the derivative of both sides of this equation and using (3.2) we find $R' = 0$. In other words, Eq. (3.38) describes a classical spacetime of constant positive curvature, i.e., de Sitter spacetime. Notice that the value of the curvature R is not fixed by this equation. For a given α, any value of $R > 12e^{-2\alpha}$ can satisfy $p = -m(1 + x^2)$. Since $\alpha \equiv t$, we arrive at the following physical interpretation of ψ: *The "wave function of the universe" $\psi(x,t)$ corresponds to a superposition of classical de Sitter spacetimes with $R > 12e^{-2t}$.* In this sense the wave function ψ can be said to describe an inflationary universe.

Since there is no length scale in this model, one could not expect to see a transition from the inflationary phase to a Robertson-Walker phase. In the future we will investigate more complicated models which include $\Lambda > 0$ and anisotropy. We hope to show that the wave function defined by imposing compact boundary conditions on the Euclidean functional integral with a positive-definite action gives a reasonable description of the observed universe.

ACKNOWLEDGMENTS

It is a pleasure to thank David Boulware, John Cardy, and especially Jim Hartle for discussions, and the Institut Henri Poincaré for its hospitality. This research was supported in part by NSF Grant No. PHY81-07384.

---

[1] J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).

[2] S. W. Hawking, Nucl. Phys. **B239**, 257 (1984); S. W. Hawking and J. C. Luttrell, Phys. Lett. **143B**, 83 (1984).

[3] S. W. Hawking and J. C. Luttrell, Nucl. Phys. **B247**, 250 (1984).

[4] See, e.g., S. W. Hawking, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, New York, 1979).

[5] For a related discussion, see A. Vilenkin, Phys. Rev. D **30**, 509 (1984).

[6] We adopt Planck units $G = c = \hbar = 1$ and the sign conventions of C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravita-*

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

tion (Freeman, San Francisco, 1973).

[7]G. W. Gibbons, S. W. Hawking, and M. J. Perry, Nucl. Phys. **B138**, 141 (1978).

[8]R. M. Schoen and S.-T. Yau, Phys. Rev. Lett. **42**, 547 (1979).

[9]T. Aubin, in *Seminar on Differential Geometry*, edited by S.-T. Yau (Princeton University Press, Princeton, New Jersey, 1982).

[10]If the negative-action conjecture is true, one might be tempted to simply change the overall sign of the action. After all, the sign of the Euclidean action was determined by "rotating" the Lorentzian action in the same direction that makes the matter action positive. But we have already seen that there are difficulties keeping both the matter and gravitational actions positive. However $-I_{\rm GR}$ would still be negative when conformal fluctuations were taken into account.

[11]See articles by D. Boulware, A. Strominger, and E. T. Tomboulis, in *Quantum Theory of Gravity*, edited by S. Christensen (Adam Hilger, Bristol, 1984); see also H. W. Hamber and R. M. Williams, Nucl. Phys. **B248**, 145 (1984).

[12]E. T. Tomboulos, Phys. Rev. Lett. **52**, 1173 (1984); D. G. Boulware, G. T. Horowitz, and A. Strominger, Phys. Rev. Lett. **50**, 1726 (1983).

[13]S. Adler, Rev. Mod. Phys. **54**, 729 (1982).

[14]F. David, Phys. Lett. **138B**, 383 (1984); F. David and A. Strominger, *ibid.* **143B**, 125 (1984).

[15]Interestingly enough, one can obtain an arbitrarily small cosmological constant by taking $\Lambda = O(\epsilon)$ and $B = O(\epsilon^{-1})$. However, it now appears to be more difficult to recover general relativity in an appropriate limit.

[16]Boulware (Ref. 11).

[17]The $D_a D_b P^{ab}$ term in (2.3) can be viewed as part of the $A_{ab}\delta/\delta K_{ab}$ term in (2.8), by multiplying the constraint by an arbitrary function and integrating by parts twice.

[18]R. Penrose, in *General Relativity: An Einstein Centenary Survey* (Ref. 4).

[19]This argument shows that no boundary terms are needed, but it does not prove that they cannot exist. Indeed, in the asymptotically flat context ($\Lambda = 0$) one expects that the Hamiltonian will contain a surface term which is just the usual Arnowitt-Deser-Misner (ADM) energy of general relativity. This suggests that the action I should contain the same boundary term as $I_{\rm GR}$. (I thank A. Ashtekar for pointing this out.) However in the cosmological context considered here, the surface term in the Hamiltonian vanishes, and we can consistently ignore surface terms in the action. For another discussion of surface terms in curvature-squared theories see N. Barth, Ph.D. thesis, University of North Carolina, 1983.

[20]G. T. Horowitz (in preparation).

[21]α' is related to the trace of the extrinsic curvature by $K = 3\alpha' e^{-\alpha}$. In Friedmann models of our universe, α' is just the Hubble "constant."

[22]For a discussion of the relation between factor ordering and path-integral measures, see K. Kuchar, J. Math. Phys. **24**, 2122 (1983), and references cited therein.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PHYSICAL REVIEW D VOLUME 31, NUMBER 12 15 JUNE 1985

Dimension of the Universe

Wu Zhong Chao
Department of Physics, Syracuse University, Syracuse, New York 13210
*and Center for Astrophysics, China University of Science and Technology, Hefei, Anhwei, China**
(Received 9 January 1985)

The dynamical evolution of the Universe is represented by the wave packet of Hartle and Hawking's "ground state." Its most probable trajectory arises by a quantum tunneling from a gravitational instanton. We discuss the dimensionality of the Universe within the framework of eleven-dimensional supergravity as a true Kaluza-Klein theory. By using some minisuperspace ansatz, it is proved that only $S_4 \times S_7$ instantons are possible. From the viewpoint of quantum cosmology, it seems that the main reason the apparent spacetime is four-dimensional is simply that there does not exist any gravitational instanton without a four-dimensional sector in the theory.

I. INTRODUCTION

Interest in explaining why spacetime is four-dimensional is by no means new. However, most of the explanations made to date essentially are based, implicitly or explicitly, on the anthropic principle.[1] The reason that the Universe looks as it does is simply a consequence of our existence. Specifically, if space is more than three-dimensional, no stable bound orbits of planets around the Sun, which are necessary for planetary life to develop, are possible in both Newton's and Einstein's gravitational theories. In quantum mechanics the same argument applies to the existence of stable atoms, which is the basis for chemistry and biology. Also, if the development of intelligence crucially depends on the information process or communication, it seems that space must have an odd number of dimensions to avoid the reverberation phenomenon.[2]

Although the anthropic principle is useful in explaining many astronomical observations such as the isotropy, homogeneity, and dimensionality of the Universe, its disadvantage is obvious: it is assumed that our inhabitable Universe is picked out from many alternative universes unsuitable for human life.

There is a common expectation in gravity theory that although spacetime appears smooth, nearly flat, and four-dimensional on large scales, at sufficiently small distances, the Planck length, it is highly curved with all possible topologies and of arbitrary dimension. The quantum state exhibits fluctuation among these possibilities. It is learned from quantum gravity[3] that this so-called spacetime foam structure has about one unit of topology per Planck volume and can be viewed as the fluctuation with chaotic creation and annihilation of mini black holes. From the discussion on spacetime foam, it turns out that the four-dimensionality of spacetime is by no means to be taken for granted. It crucially depends on how the initial conditions are chosen for the Universe. We are brought face-to-face with a dilemma: the initial conditions for the evolution of the Universe must be imposed by hand, but the Universe must be self-contained, since there is nothing outside of it.

Great progress in understanding the origin of the Universe has been made by the new developments in quantum cosmology.[4] The long-standing "first cause" problem intrinsic in cosmology is dispelled by the ground-state proposal of Hartle and Hawking. In the big-bang cosmological model, in the very early stage of the Planck era, quantum gravitational effects played the most important role. The quantum state of the Universe is defined by the ground state of Hartle and Hawking. The classical evolution trajectories are represented by the wave packet of the wave function. Recall that before the emergence of the ground-state proposal in quantum cosmology the initial conditions needed for the field evolution equations were inserted by hand, so all evolutions but that represented by the wave packet are not relevant to reality. Furthermore, among these possible trajectories of the wave packet one can single out the most probable one which is supposed to describe the universe we live in.

Nearly all investigations on the origin of spacetime dimensions are done within the framework of Kaluza-Klein models.[5] It is believed that the extra dimensions have been spontaneously compactified into the Planck length scale. However, not until the discovery of simple supergravity in $d=11$ dimensions[6] did we know which Kaluza-Klein theory was probably a realistic one. And only after the presentation of an acceptable ground-state proposal will we know what boundary conditions must be imposed on the Kaluza-Klein theory.

In this paper we shall discuss the dimensionality of the Universe from the viewpoint of quantum cosmology. By using $d=11$ simple supergravity as a true Kaluza-Klein theory, we can show how the dimension of the Universe came from the ground state of Hartle and Hawking. Section II gives a brief review of the Hartle and Hawking proposal and generalizes it to the cases of some noncompact or higher-dimensional spacelike geometries as the configuration space of the wave function. A five-dimensional Kaluza-Klein model is presented as a toy model.[7] In Sec. III we show that the most probable evolution must arise by a quantum tunneling from gravitational instantons. In a higher-dimensional model with a positive cosmological constant, the Universe most probably

 ©1985 The American Physical Society

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

evolves with a static internal space of a given constant size.[7]

In Sec. IV we turn to the realistic $d=11$ supergravity model and prove that under the minisuperspace ansatz with topology $S_k \times S_{11-k}$, only $S_4 \times S_7$ instantons are possible. The gauge potential of these instantons on Einstein space S_7, up a gauge transformation, must be proportional to the Ricci-flattening torsion tensor on S_7. Perhaps, the instanton versions of the Freund-Rubin,[8] Duff-Pope,[9] Englert,[10] and Englert-Rooman-Spindel[11] solutions exhaust the minisuperspace solutions under the $S_4 \times S_7$ ansatz. By extending our discussion to the case of $S_m \times S_n \times \cdots \times S_k$ $(m+n+\cdots+k=11)$ topology, we find that for any nontrivial instanton, an S_4 sector is not avoidable. A conjecture is as follows: for the most general ansatz any instanton must have a four-dimensional sector. The nonexistence of instantons without a four-dimensional sector may be the main reason why spacetime is apparently four-dimensional. However, we do not think this conjecture will be proved in the near future.

II. THE GROUND STATE OF THE UNIVERSE

The quantum state of the Universe is described by the ground state of Hartle and Hawking,[4]

$$\Psi[h_{ij},\phi_0]=\int_C \delta[g_{\mu\nu}]\delta[\phi]\exp(-\bar{I}[g_{\mu\nu},\phi]) , \tag{2.1}$$

where the wave function has compact spacelike three-geometry h_{ij} and matter field ϕ_0 on it as its configuration. In Euclidean quantum gravity the path integral is summed over all compact positive-definite four-metrics and matter fields on them which have the configuration (h_{ij},ϕ_0) as a boundary [Fig. 1(a)]. $\bar{I}$ is the Euclidean action.

Since the quantum state does not depend on the time label and choice of three spatial coordinates on the three-geometry, the wave function must obey the Wheeler-DeWitt equation[12]

$$\left\{-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}} + h^{1/2}\left[-{}^3R+16\pi T_\infty\left[\frac{\delta}{\delta\phi},\phi\right]\right]\right\}\Psi[h_{ij},\phi]=0 , \tag{2.2}$$

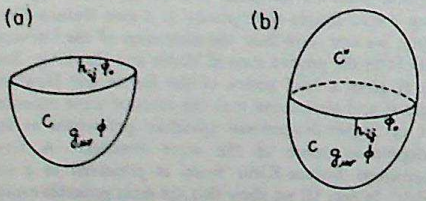


FIG. 1. (a) The ground state of the Universe is defined by a path integral from nothing in Euclidean quantum gravity. (b) The probability for tunneling to the Lorentzian evolution of the Universe is formulated by a path integral from nothing to nothing in Euclidean quantum gravity.

where

$$G_{ijkl}=\tfrac{1}{2}h^{-1/2}(h_{ik}h_{jl}+h_{il}h_{jk}-h_{ij}h_{kl}) , \tag{2.3}$$

and the momentum constraints

$$\left[\frac{\delta\Psi}{\delta h_{ij}}\right]_{|i}=8\pi T^{0}{}_{j}\Psi . \tag{2.4}$$

In general, the configuration space of the ground state contains some so-called Lorentzian region[4] where the wave function behaves like a wave packet

$$\Psi=C\cos S , \tag{2.5}$$

where C is the slowly changing amplitude and S the rapidly varying phase. In the WKB approximation, the Wheeler-DeWitt equation for the wave packet can be reduced to the Hamilton-Jacobi equation with S identified as the classical action. Thus the wave packet can be interpreted as an ensemble of classical Lorentzian trajectories which satisfy the equation

$$\nabla_i S=P_i , \tag{2.6}$$

where P_i are the canonical momenta.

The ground-state proposal gives very strong restriction to classical evolutions. Indeed, the second-order Einstein equation has been truncated into the first-order equation (2.6) by the proposal. In a minisuperspace model with n variables, the general evolution has $2n-2l$ number degrees of freedom, where l denotes the number of the Einstein and momentum constraints, since these constraints represent the coordinate (or gauge) invariance of the system, the gauge-independent degrees of freedom must be reduced to $2n-2l$. This number is consistent with the consideration in the ordinary canonical mechanics, the action takes both ends of trajectories as the configuration. While for the phase S of the wave packet, one can view the south pole of the four-metrics as an end point. Thus, the number of degrees of freedom of the evolution is reduced to $n-l$, or, by half. Therefore, in the oscillating region, in general, there is one and only one trajectory through each point.

The region where the wave function does not take wave-packet form is called the Euclidean region. In this region the time concept breaks down and the classical evolution becomes meaningless.

The ground-state proposal is very powerful in predicting the whole Universe solely from known physical laws. Every model of quantum cosmology seems a step toward the realistic universe.[4,7,13] Unfortunately, this proposal can only be applied to the case of compact three-geometry as the configuration. Recently, Hartle[14] has generalized it to the case of asymptotically flat three-geometry with matter on it. Then the class C of the path integral are all asymptotically flat four-manifolds with matter on them. It turns out that for a flat spacetime background this definition of ground state is consistent with the minimum-energy state in quantum field theory.

The proposal of Hartle and Hawking can be generalized to the case of any three-geometry in which the complexified three-manifold has a real compact section.[15] Then one can take the path integral for the ground state over all

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

compact four-manifolds with this section as a boundary. After evaluating the path integral, one can get the wave function for the original three-geometry as the configuration by an analytic continuation.[7] If the complexified three-manifold has many real compact sections, the quantum state of the Universe becomes degenerate and it is the mixture of ground states defined by these sections. This phenomenon is similar to the θ vacuum in quantum Yang-Mills theory. The restriction for this generalized proposal is that the subset of having compact section is only of measure zero among all general noncompact three-geometries. A general proposal for any noncompact three-geometry as configuration is still not available. Fortunately, it does not affect our discussion on the minisuperspace models below. In general, in the case of Hartle and Hawking, the classical trajectories of the wave packet can be thought of as being created by quantum tunneling from compact Euclidean manifolds. The quantum tunneling can be visualized most clearly in the Euclidean section with an imaginary time. However, there does exist an alternative phase transition, i.e., particle creation and annihilation, and one must visualize it in a spacetime section with imaginary spatial coordinates. Physically, this is exactly what we do for our generalized proposal.[7]

For example, in the anti–de Sitter model, the spatial three-hyperboloid can be compactified from

$$ds^2=-dt^2+a^2(t)[dx^2+\sinh^2x(d\theta^2+\sin^2\theta\,d\phi^2)] \tag{2.7}$$

to

$$ds^2=-dt^2-a^2(t)[d\kappa^2+\sin^2\kappa(d\theta^2+\sin^2\theta\,d\phi^2)] \tag{2.8}$$

by the substitution $x\rightarrow\pm i\kappa\pm in\pi$ $(n=0,1,2,\ldots)$. The main contribution to the ground-state wave function comes from the Euclidean solution S_4 of radius $H^{-1}\equiv(3/|\Lambda|)^{1/2}$, where Λ is a negative cosmological constant, and the wave function at the WKB level becomes

$$\Psi(a)=N_0\cos\left[\frac{1}{3H^2}[1-(1-H^2a^2)^{3/2}]\right] \quad (Ha<1)\,, \tag{2.9}$$

$$\Psi(a)=N_1\cos\left[\frac{\pi}{4}+\frac{1}{3H^2}\right]\exp[-(H^2a^2-1)^{3/2}/3H^2] \quad (Ha>1)$$

$$(N_0,N_1 \text{ const})\,.$$

Of course one can derive the anti–de Sitter metric from the wave function for $Ha<1$ by using correspondence (2.6).

The ground-state proposal can also be applied to higher-dimensional problems simply by increasing the relevant dimensions in the formulation. The simplest model is the Kaluza-Klein cosmology with a positive cosmological constant and the metric form

$$ds^2=-dt^2+a^2(t)d\Omega_3{}^2+b^2(t)d\Omega_1{}^2\,, \tag{2.10}$$

where $d\Omega_n{}^2$ denotes the unit n-sphere metric. All classical Lorentzian and Euclidean solutions are known.[16] However, only the following Euclidean solutions are compact and regular at the south pole:

$$ds^2=d\tau^2+\frac{6}{\Lambda}\sin^2\left[\left(\frac{\Lambda}{6}\right)^{1/2}\tau\right]d\Omega_3{}^2 +C_3{}^2\cos^2\left[\left(\frac{\Lambda}{6}\right)^{1/2}\tau\right]d\Omega_1{}^2 \quad (C_3 \text{ const})\,. \tag{2.11}$$

In the semiclassical approximation, the path integral (2.1) can be written as

$$\Psi=N_0\sum_i A_i\exp(-B_i)\,, \tag{2.12}$$

where N_0 is the normalization factor, B_i is the Euclidean actions for the classical orbits, and A_i represents the fluctuations around them. At the same level we obtain the wave function for $a<(\Lambda/6)^{-1/2}$,

$$\Psi(a,b)=N_0\exp[\tfrac{1}{4}a^2b(1-H^2a^2)^{1/2}] \quad [N_0 \text{ const},\ H^{-1}\equiv(\Lambda/6)^{-1/2}]\,. \tag{2.13}$$

The wave function for $a>H^{-1}$ can be obtained by an analytic continuation

$$\Psi(a,b)=N_0'\cos[\tfrac{1}{4}a^2b(H^2a^2-1)^{1/2}+\delta] \quad (N_0',\delta \text{ const})\,. \tag{2.14}$$

The classical trajectories of the wave packet (2.14) are found by substituting (2.14) into (2.6),

$$ds^2=-dt^2+\frac{6}{\Lambda}\cosh^2(\sqrt{\Lambda/6}t)d\Omega_3{}^2 +C_3{}^2\sinh^2(\sqrt{\Lambda/6}t)d\Omega_1{}^2 \quad (C_3 \text{ const})\,. \tag{2.15}$$

It follows that the circle of the fifth dimension eventually expands at the same rate as the three-sphere and the spontaneous compactification does not occur.

III. THE MOST PROBABLE EVOLUTION

The simple five-dimensional Kaluza-Klein model does not give rise to spontaneous compactification. Let us consider the next model with a positive cosmological constant and $R\times S_3\times S_n$ $(n\geq 2)$ metric form

$$ds^2=-dt^2+a^2(t)d\Omega_3{}^2+b^2(t)d\Omega_n{}^2\,. \tag{3.1}$$

Its gravitational action can be written as

$$I=\int dt\,R^{n+3}\left[(n+2)(n+3)\left[\frac{-\dot R^2}{R^2}+\frac{3n\dot r^2}{(n+2)r^2}\right]+\frac{1}{R^2}[6r^{-2n}+(n-1)nr^6]-2\Lambda\right]\,, \tag{3.2}$$

where we have introduced the scale and anisotropy parameters

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$a = Rr^n, \quad b = Rr^{-3} . \tag{3.3}$$

As usual, the Wheeler-DeWitt equation is hyperbolic with the scale parameter R as a time coordinate,

$$\left[-\frac{1}{R^p}\frac{\partial}{\partial R}R^p\frac{\partial}{\partial R} + \frac{\partial^2}{R^2\partial\rho^2} + \frac{R^{2n+2}}{(n+2)(n+3)}\left[6e^{-2n\rho[(n+2)/3n]^{1/2}} + n(n-1)e^{6\rho[(n+2)/3n]^{1/2}} - 2\Lambda R^2\right]\right]\Psi(R,\rho) = 0 ,$$

$$\rho \equiv \left[\frac{3n}{n+2}\right]^{1/2} \ln r . \tag{3.4}$$

The index p represents some, but not all, of the ambiguity in the operator ordering or, equivalently, the measure in the path integral. By setting $p=1$ and making a coordinate transformation again

$$x = R\sinh\rho, \quad y = R\cosh\rho , \tag{3.5}$$

one obtains a standard hyperbolic equation with the causal future of the origin as the physically feasible region. We have only two obvious Euclidean solutions:

(1) $S_4 \times S_n$

$$a = \left[\frac{3(n+2)}{2\Lambda}\right]^{1/2}\cos\left[\left[\frac{2\Lambda}{3(n+2)}\right]^{1/2}\tau\right], \quad b = b_0 \equiv \left[\frac{(n+2)(n-1)}{2\Lambda}\right]^{1/2} , \tag{3.6}$$

(2) $S_3 \times S_{n+1}$

$$a = a_0 \equiv \left[\frac{n+2}{\Lambda}\right]^{1/2}, \quad b = \left[\frac{n(n+2)}{2\Lambda}\right]^{1/2}\cos\left[\left[\frac{2\Lambda}{n(n+2)}\right]^{1/2}\tau\right] . \tag{3.7}$$

In order to obtain the wave function, one has to numerically integrate the Wheeler-DeWitt equation, instead of using the WKB approximation. Recall that the ground state is defined by a path integral over all compact and regular metrics, the action for Euclidean metrics at the points near the null cone $R=0$ is very small, and for our choice $p=1$, it is consistent to take the boundary condition there to be $\Psi=1$ and its first-order derivative to be zero. The $\Psi=1$ condition at $x=y$ comes from the Euclidean metrics with regularity condition at the south pole $b=0$, $db/d\tau=1$, $da/d\tau=0$, while at $x=-y$ the condition should be $a=0$, $da/d\tau=1$, and $db/d\tau=0$.

Before R reaches the zero-potential surface from below

$$\frac{6}{a^2} + \frac{n(n-1)}{b^2} - 2\Lambda = 0 , \tag{3.8}$$

the wave function takes an exponential behavior. It starts to oscillate at the spacelike (with respect to the supermetric) segment of this surface

$$\frac{3n-(3n^2+6n)^{1/2}}{6} < \left[\frac{b}{a}\right]^2 < \frac{3n+(3n^2+6n)^{1/2}}{6} . \tag{3.9}$$

Thus, all classical evolutions must start from this segment with zero initial velocity. The condition of zero initial velocity comes from the Einstein constraint, the classical counterpart of the Wheeler-DeWitt equation. Among these trajectories, two obvious exact ones are the Lorentzian versions of metrics (3.6) and (3.7):

(1) $$a = \left[\frac{3(n+2)}{2\Lambda}\right]^{1/2}\cosh\left[\left[\frac{2\Lambda}{3(n+2)}\right]^{1/2}t\right],$$

$$b = b_0 , \tag{3.10}$$

(2) $$b = \left[\frac{n(n+2)}{2\Lambda}\right]^{1/2}\cosh\left[\left[\frac{2\Lambda}{n(n+2)}\right]^{1/2}t\right],$$

$$a = a_0 . \tag{3.11}$$

They have an internal S_3 or S_n space with given constant sizes a_0, b_0. If the initial S_3 size is smaller than a_0, then it will recollapse, in general, at Planck time scale while the external S_n will expand to infinity. If the initial size of b is smaller than the critical value b_0, their behaviors are interchanged. For the case with both initial sizes bigger than the critical values, both S_3 and S_n will expand forever.

It is expected that near the recollapse of the internal space, the quantum gravitational effect plays an important role. In fact, the semiclassical approximation would break down there. Not only will expression (2.12) fail to give a reliable wave function, but also one cannot reduce the Wheeler-DeWitt equation into the Hamilton-Jacobi equation by simply throwing away the higher-order correction terms. Instead one must put these corrections at the right-hand side of the equation. We expect to see the detail of the evolution near the recollapse stage from the wave function obtained by numerical integration.

One can view the transition from the Euclidean region to the Lorentzian region at zero-potential surface as a form of quantum tunneling. The relative probability for a tunneling through the configuration (a,b) at the zero-potential surface can be formulated by a path integral from nothing to nothing [Fig. 1(b)],

$$\Psi^*\Psi = \int_{C+C^*}\delta[g_{\mu\nu}]\delta[\phi]\exp(-\bar{I}[g_{\mu\nu},\phi]) , \tag{3.12}$$

where the class $C+C^*$ are all compact positive-definite metrics which have the configuration (a,b) as a section.

The instanton solution of the Einstein equation dom-

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

inates the path integral, i.e., if $C+C^*$ is an instanton, then the probability takes an extreme value.

For our concrete model, metrics (3.6) and (3.7) are the only instantons. Thus, their Lorentzian versions (3.10) and (3.11) describe the most probable evolutions of the Universe. For the general case, the regularity condition at the north pole of instantons is as strong as that at the south pole, therefore, it singles out the most probable evolution trajectory.

It must be emphasized that the topology of S_3 and S_n are not essential for this model, but the curvature terms are very important. In the absence of these terms, the qualitative behavior of the evolution is rather different, even at the classical level. For example, all exact solutions called Kasner–de Sitter metrics[17] which are found for flat-spacelike-sector cases show that all sectors will eventually expand exponentially at the same rate.

The models discussed previously are not realistic, since the dimension and the cosmological constant are arbitrarily put in by hand. Many people consider $d=11$ supergravity as a true Kaluza-Klein theory, so we shall discuss it in the next section.

IV. INSTANTONS IN $d=11$ SUPERGRAVITY

Kaluza-Klein theory with $d=11$ simple supergravity is a more realistic model. The eleven-dimensional spacetime perhaps is realistic and by the spontaneous compactification the extra seven dimensions are squeezed into the Planck length scale, representing the internal symmetry of particle physics.[18] Supersymmetry compels us to introduce a spin-$\frac{3}{2}$ gravitino field and a three-index antisymmetric tensor A_{MNP} into the theory. For simplicity, we assume the gravitino field to be zero below. Then the field of equations are

$$R_{MN}-\tfrac{1}{2}g_{MN}R=\tfrac{1}{48}(8F_{MPQR}F_{N}{}^{PQR}-g_{MN}F_{SPQR}F^{SPQR}), \tag{4.1}$$

$$F^{MNPQ}{}_{;M}=\left|\frac{-\sqrt{2}}{2(4!)^2}\right|\eta^{M_1\cdots M_8NPQ}F_{M_1\cdots M_4}F_{M_5\cdots M_8} \qquad (0\le M,N,P,Q,S,R,\ldots\le 10), \tag{4.2}$$

where $F_{MNPQ}=4!\partial_{[M}A_{NPQ]}$ and $\eta^{A\cdots N}=|g|^{-1/2}\times\epsilon^{A\cdots N}$. In the following we shall work with a minisuperspace model of a finite number of degrees of freedom. It is assumed that the spacetime takes a product form of Einstein spaces $M_k\times M_{11-k}$ and the only nonvanishing field components are

$$g_{mn}(x),\ g_{\mu\nu}(y)\ (m,n,p,q=0,\ldots,k-1),$$

$$F_{mnpq}(x),\ F_{\mu\nu\rho\sigma}(y)\ (\mu,\nu,\rho,\sigma=k,\ldots,10).$$

One is tempted to find the wave function of the ground state, specifying M_k to be $R\times S_{k-1}$ and M_{11-k} to be S_{11-k}, where R denotes time, and S_{k-1} and S_{11-k} are interpreted as the external and internal spaces, respectively. The Wheeler-DeWitt equation is hyperbolic with the geometric average of the scale lengths ω as the only timelike coordinate. In the configuration space, when ω is sufficiently small, the action for Euclidean metrics is very small. The wave function takes some polynomial or exponential form there, its exact form depends on the operator ordering in the Wheeler-DeWitt equation. One expects that as soon as it reaches a spacelike segment of the zero-potential surface, the wave function would start to oscillate and the Universe would begin to evolve classically. Unfortunately, the zero-potential surface remains timelike everywhere and no Lorentzian region exists. This corresponds to the fact that there does not exist any classical Lorentzian solution under this ansatz.

However, if we assume $M_k=R\times H_{k-1}$ ($k\ge 4$) (H_m is an m-dimensional hyperboloid) and $M_{11-k}=S_{11-k}$, or $M_k=R\times S_{k-1}$ ($k=2,3$) and $M_{11-k}=H_{11-k}$, the situation changes drastically. One can find the classical solution with topology $R\times S_{k-1}\times S_{11-k}$ and of metric signature

$$(\underbrace{-\cdots-}_{k}\ \underbrace{+\cdots+}_{11-k})$$

and

$$(\underbrace{+\cdots+}_{k}\ \underbrace{-\cdots-}_{11-k}),$$

respectively. (S_m is considered as a compact real sector of complexified H_m.) These solutions dominate the path integral for the ground state. Returning to the H_{k-1} or H_{11-k} configuration the wave function oscillates and represents the classical Lorentzian orbits of the ground state. Possible orbits exist for $k=2,3,\ldots,9$; it means that one can have a macroscopic spacetime of $2,3,\ldots,9$ dimensions. Considering dimensionality, one hopes maybe the most probable evolution of the Universe must have $k=4$. This is indeed the case.

In order to obtain the most probable evolution, one has to find the instanton solutions. Now we are looking for instantons with topology $S_k\times S_{11-k}$. If $k=1,2,3$, we have $F_{mnpq}=0$ and Eq. (4.2) implies that

$$F^{\mu\nu\rho\sigma}{}_{;\sigma}=0 \tag{4.3}$$

or, equivalently,

$$d^*F=0, \tag{4.4}$$

where $^*F$ is the Hodge duality of F in S_{11-k} space. Combining with the fact F is the gauge field of potential A, it turns out that F must be a harmonic four-form on S_{11-k}. Noting that the four-dimensional de Rham cohomology vector space $H^4(S_{11-k})$ is zero dimensional,[19] we obtain

$$F_{\mu\nu\rho\sigma}(y)=0. \tag{4.5}$$

For $k=5$, the vanishing of F_{mnpq} components is implied by the fact that $H^4(S_5)$ is zero and by Eq. (4.2). And by exactly the same method, we find condition (4.5). It is worth mentioning that the regularity condition at the "north pole" of the S_5 sector of the instanton is important. In the absence of this condition, for example, the noncompact Lorentzian version of S_5 will have a nonvanishing harmonic form F_{mnpq}.

Of course, all these arguments apply also to $k=6,8,9,10$ by dimension duality. It is concluded that the only possibility to find the instanton solution is the $S_4\times S_7$ case.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

For the $S_4 \times S_7$ case, because $H^4(S_4)=1$, the gauge field strength F_{mnpq} must take the harmonic form

$$F_{mnpq} = f\eta_{mnpq}, \quad f = \text{const}. \tag{4.6}$$

If we assume

$$F_{\mu\nu\rho\sigma} = 0, \tag{4.7}$$

then from the Einstein equation we conclude that the Lorentzian version of the spacetime must be either a Freund-Rubin space which is a product of an anti–de Sitter space and a round S_7, $[SO(3,2)/SO(3,1)] \times SO(8)/SO(7)$, or a Duff-Pope space in which S_7 is squashed and given by the Einstein metric on the coset space $[SP(2)\times SP(1)]/SP(1)\times SP(1)$. The former solution gives rise to an effective four-dimensional model with local SO(8) invariance and $N=8$ supersymmetry, while the latter case gives only $SO(5)\times SU(2)$ invariance and $N=1,0$ supersymmetry. The effective cosmological constants for the external and internal spaces are $-4f^2/3$ and $2f^2/3$, respectively, and the instanton itself has metric signature $(-----+++++++)$.

If the S_7 section of the gauge-field strength does not vanish, it can be written

$$F_{\mu\nu\rho\sigma} = \frac{4}{f} S_{[\mu\nu,\sigma]}. \tag{4.8}$$

For round S_7, Englert[10] found that if $S_{\mu\nu\rho}$ is one of the two totally antisymmetric torsion tensor $S_{\pm\mu\nu\rho}$ which flatten the S_7 in the sense that

$$R^{\mu}_{\nu\rho\sigma}(\Gamma^{\tau}_{\tau\omega} + S^{\tau}_{\tau\omega}) = 0, \tag{4.9}$$

then condition (4.8) solves the S_7 section of field equation (4.2),

$$F^{\mu\nu\rho\sigma}{}_{;\mu} = \frac{\sqrt{2}}{4!} f \eta^{\nu\rho\sigma\tau\omega\delta} F_{\tau\omega\delta}. \tag{4.10}$$

For squashed S_7, Englert, Rooman, and Spindel[11] found solutions with torsion tensors $S_{\pm\mu\nu\rho}$ which render S_7 Ricci-flat,

$$R_{\mu\nu}(\Gamma^{\tau}_{\tau\omega} + S^{\tau}_{\tau\omega}) = 0. \tag{4.11}$$

The gauge field of forms (4.7) and (4.8) results in an anisotropic cosmological constant $(-5f^2/2, 3f^2/2)$. No supersymmetry is preserved for these two cases.

In fact, we can prove that all nontrivial solutions of (4.10) must take the form (4.8) with $S_{\mu\nu\rho}$ proportional to a Ricci-flattening torsion tensor. For any solution of (4.10) we can define

$$T_{\mu\nu\rho} = K\eta_{\mu\nu\rho\sigma\tau\omega\delta} F^{\sigma\tau\omega\delta} \quad (K \text{ const}). \tag{4.12}$$

The Ricci tensor with the torsion $T_{\mu\nu\rho}$ takes the value

$$R_{\sigma\delta}(\Gamma^{\mu}_{\nu\rho} + T^{\mu}_{\nu\rho}) = R_{\sigma\delta}(\Gamma^{\mu}_{\nu\rho}) - T_{\tau\sigma\omega} T^{\omega\tau}{}_{\delta}. \tag{4.13}$$

One can always choose an appropriate coefficient K to satisfy the condition $R_{\sigma\delta}(\Gamma^{\mu}_{\nu\rho} + T^{\mu}_{\nu\rho}) = 0$. For the Englert and Englert, and Rooman and Spindel solutions we have $K = 1/\sqrt{24!}$. It follows from (4.10) and (4.12) that $d^*F(y) \propto F(y)$ and $T(y) \propto {}^*F(y)$, thus we have $F(y) \propto dT(y)$.

We have listed all known instanton solutions of the topology $S_4 \times S_7$. The significance of the Freund-Rubin solution is that by assuming the gauge field to live on M_4, S_7 compactification occurs spontaneously. We show that, in quantum cosmology, only instanton solutions are important and for instanton solutions this phenomenon in fact is implied by the theory itself. It was very surprising at first that the Englert solution was obtained by relating the gauge field in S_7 with its flattening torsion tensor. Now it is realized that for all solutions the gauge potential on Einstein space S_7 must be proportional to a Ricci-flattening on the round or squashed S_7. However, at this moment, one does not know whether these four cases exhaust the instanton solutions to the field equations.

V. DISCUSSION

The quantum state of the Universe must be defined by the Hartle-Hawking state which represents an ensemble of classical trajectories for the evolution of the Universe. The instanton solution dominates the quantum tunneling probability and singles out the most probable evolution.

We have proved that in eleven-dimensional supergravity, under the minisuperspace ansatz $S_k \times S_{11-k}$, the only nontrivial instanton takes a $S_4 \times S_7$ form. Returning to the realistic Lorentzian spacetime, this fact implies that the spontaneous compactification in $d=11$ supergravity restricts our observed spacetime to be four- or seven-dimensional, and by causality condition, there is only one time coordinate. However, the dimension seven is not essential. In fact, in the Freund-Rubin instanton, one can easily replace the S_7 sector by $S_2 \times S_2 \times S_3$, $S_3 \times S_4$, or $S_2 \times S_5$ sectors. And by the same argument used for $S_k \times S_{11-k}$, one can prove that for the ansatz $S_m \times S_n \times \cdots \times S_k$ $(m+n+\cdots k=11)$ the S_4 sector is not avoidable.

All known $S_4 \times S_7$ instanton solutions suffer from the common difficulty that the observed spacetime must share a comparable cosmological constant with the internal space of Planck length scale, this fact simply conflicts with the observation: the apparent cosmological constant is extremely small or probably zero. It seems that to account for the cosmological constant problem, one has to go beyond the minisuperspace model[20] or to include the contributions of fermionic fields.

The ansatz we used is rather special. The reason is simply that it makes the problem tractable. We expect in $d=11$ simple supergravity, any instanton must have a four-dimensional sector and this is the main reason why spacetime is apparently four-dimensional. Nevertheless, we do not think this conjecture will be proved in the near future.

ACKNOWLEDGMENTS

I would like to thank S. W. Hawking for suggestions and J. Ehlers of the Max-Planck-Institut für Astrophysik, R. Ruffini of Rome University, A. Salam of the International Center for Theoretical Physics, Trieste, and H. Sato of Kyoto University for their hospitality. This work has been partially supported by the National Science Foundation under Grant No. PHY-8318350.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

*Permanent address.
[1]C. B. Collins and S. W. Hawking, Astrophys. J. 180, 317 (1973).
[2]J. D. Barrow, Philos. Trans. R. Soc. London A310, 337 (1983).
[3]S. W. Hawking, Nucl. Phys. B144, 349 (1978).
[4]S. W. Hawking, Pontif. Accad. Sci. Varia 48, 563 (1982); J. B. Hartle and S. W. Hawking, Phys. Rev. D 28, 2960 (1983); S. W. Hawking, in *Relativity, Groups and Topology II*, Les Houches Session XL, edited by B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).
[5]T. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. 966 (1921); O. Klein, Z. Phys. 37, 895 (1926).
[6]E. Cremmer, B. Julia, and J. Scherk, Phys. Lett. 76B, 409 (1978).
[7]Wu Zhong Chao, Phys. Lett. 146B, 307 (1984); Hu Xiao Ming and Wu Zhong Chao, *ibid.* 149B, 87 (1984); Hu Xiao Ming and Wu Zhong Chao, *ibid.* (to be published).
[8]P. G. O. Freund and M. A. Rubin, Phys. Lett. 97B, 233 (1980).
[9]M. J. Duff and C. N. Pope, in *Supersymmetry and Supergravity 82*, edited by S. Ferrara, J. G. Taylor, and P. Van Nieuwenhuizen (World Scientific, Singapore, 1983).
[10]F. Englert, Phys. Lett. 119B, 339 (1982).
[11]F. Englert, M. Rooman, and P. Spindel, Phys. Lett. 127B, 47 (1983).
[12]B. S. DeWitt, Phys. Rev. 160, 1113 (1967).
[13]S. W. Hawking, Nucl. Phys. B239, 257 (1984); S. W. Hawking and J. C. Luttrell, Phys. Lett. 143B, 83 (1984); S. W. Hawking and J. C. Luttrell, Nucl. Phys. B247, 250 (1984); S. W. Hawking and Z. C. Wu, Phys. Lett. 151B, 15 (1985).
[14]J. B. Hartle, Phys. Rev. D 29, 2730 (1984).
[15]S. W. Hawking, Phys. Lett. 60A, 81 (1977).
[16]H. Ishihara, K. Tomita, and H. Nariai, Prog. Theor. Phys. 71, 859 (1984).
[17]H. Sato, Prog. Theor. Phys. 72, 98 (1984).
[18]A. Salam and J. Strathdee, Ann. Phys. (N.Y.) 141, 316 (1982).
[19]T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. 66, 213 (1980).
[20]P. Candelas and S. Weinberg, Nucl. Phys. B237, 397 (1984); P. Candelas and D. J. Raine, *ibid.* B243, 157 (1984).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

PHYSICAL REVIEW D VOLUME 32, NUMBER 10 15 NOVEMBER 1985

Arrow of time in cosmology

S. W. Hawking
University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, England
(Received 29 April 1985)

The usual proof of the *CPT* theorem does not apply to theories which include the gravitational field. Nevertheless, it is shown that *CPT* invariance still holds in these cases provided that, as has recently been proposed, the quantum state of the Universe is defined by a path integral over metrics that are compact without boundary. The observed asymmetry or arrow of time defined by the direction of time in which entropy increases is shown to be related to the cosmological arrow of time defined by the direction of time in which the Universe is expanding. It arises because in the proposed quantum state the Universe would have been smooth and homogeneous when it was small but irregular and inhomogeneous when it was large. The thermodynamic arrow would reverse during a contracting phase of the Universe or inside black holes. Possible observational tests of this prediction are discussed.

I. INTRODUCTION

Physics is time symmetric. More accurately, it can be shown[1] that any quantum field theory that has (a) Lorentz invariance, (b) positive energy, and (c) local causality, i.e., $\phi(x)$ and $\phi(y)$ commute (or anticommute) if x and y are spacelike separated, is invariant under *CPT* where *C* means interchange particles with antiparticles, *P* means replace left hand by right hand, and *T* means reverse the direction of motion of all particles. In most situations, the effect of any *C* or *P* noninvariance can be neglected, so that the interactions ought to be invariant under *T* alone.

In fact, if one takes the gravitational field into account, the Universe that we live in does not satisfy any of the three conditions listed above. The Universe is not Lorentz invariant because spacetime is not flat, or even asymptotically flat. The energy density is not positive definite because gravitational potential energy is negative. In a certain sense the total energy of the Universe is zero because the positive energy of the matter is exactly compensated by the negative gravitational potential energy. Finally, the concept of local causality ceases to be well defined if the spacetime metric itself is quantized because one cannot tell if x and y are spacelike separated. Nevertheless, I shall show in Sec. III of this paper that the universe is invariant under *CPT* if, as has been recently proposed,[2–4] it is in the quantum state defined by a path integral over compact four-metrics without boundary. This is a nontrivial result because an arbitrary quantum state for the Universe is not, in general, invariant under *CPT*.

The Universe that we live in certainly does not appear time symmetric, as anyone who has watched a movie being shown backward can testify: one sees events that are never witnessed in ordinary life, like pieces of a cup gathering themselves together off the floor and jumping back onto a table. One can distinguish a number of different "arrows of time" that express the time asymmetry of the Universe. (1) The thermodynamic arrow: the direction of time in which entropy increases. (2) The electrodynamic arrow: the fact that one uses retarded solutions of the field equations rather than advanced ones. (3) The psychological arrow: the fact that we remember events in the past but not in the future. (4) The cosmological arrow: the direction in time in which the universe is expanding.

I shall take the point of view that the first arrow implies the second and third. In the case of the psychological arrow this follows because human beings (or computers, which are easier to talk about) are governed by the thermodynamic arrow, like everything else in the Universe. In the case of electrodynamics, one can express the vector potential $A_\mu(x)$ as a sum of a contribution from sources in the past of x plus a surface integral at past infinity. One can also express $A_\mu(x)$ as a sum of a contribution from sources in the future of x plus a surface integral at future infinity. The boundary conditions that give rise to the thermodynamic arrow imply that there is no incoming radiation in the past. Thus the surface integral in the past is zero and the electromagnetic field can be expressed as an integral over sources in the past. On the other hand, the boundary conditions that give rise to the thermodynamic arrow do not prevent the possibility of outgoing radiation in the future. This means that the surface integral in the future is strongly correlated with the contribution from sources in the future. It therefore cannot be neglected.

The accepted explanation for the thermodynamic arrow of time is that for some reason the Universe started out in a state of high order or low entropy. Such states occupy only a very small fraction of the volume of phase space accessible to the Universe. As the Universe evolves in time it will tend to move around phase space ergodically. At a later time therefore there is a high probability that the Universe will be found in a state of disorder or higher entropy because such states occupy most of phase space. Consider, for example, a system consisting of a number N of gas molecules in a rectangular box which is divided

 ©1985 The American Physical Society

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

into two by a partition with a small hole in it. Suppose that at some initial time, say 10 o'clock, all the molecules are in the left-hand side of the box. Such configurations occupy only one part in 2^N of the available $6N$-dimensional phase space. As time goes on, the system will move around phase space on a constant-energy surface. At a later time there will be a high probability of finding the system in a more disordered state with molecules in both halves of the box. Thus entropy will increase with time. Of course, if one waits long enough, one will eventually see all the molecules returning to one half of the box. However, for macroscopic values of N, the time taken is likely to be much longer than the age of the Universe.

Suppose, on the other hand, that the Universe satisfied a *final* condition that was in a state of high order. In that case it would be likely to be in a more disordered state at earlier times and entropy would decrease with time. However, as remarked above, the psychological arrow is determined by the thermodynamic arrow. Thus, if the thermodynamic arrow were reversed, the psychological arrow would be reversed as well: we would define time to run in the other direction and we would still say that entropy increased with time. However, the cosmological arrow provides an independent definition of the direction of time with which we can compare the thermodynamic, psychological, and electrodynamic arrows. In the early 1960s Hogarth[5] and Hoyle and Narlikar[6] tried to connect the electrodynamic and cosmological arrows using the Wheeler-Feynman[7] direct-particle-interaction formulation of electrodynamics. At a summer school held[8] at Cornell in 1963 their work was criticized by a Mr. X (generally assumed to be Richard Feynman) on the grounds that they had implicitly assumed the thermodynamic arrow. They also got the "wrong" answer in that they predicted retarded potentials in a steady-state universe but advanced ones in an evolutionary universe without continual creation of matter. It is now generally accepted that we live in an evolutionary universe.

Another proposal to explain the thermodynamic arrow of time has been put forward by Penrose.[9] It is based on the prediction of classical general relativity[10] that there will be spacetime singularities both in the past, at the big bang, and in the future at the big crunch, if the whole universe recollapses, or in black holes if only local regions collapse. Penrose's proposal is that the Weyl tensor should be zero at singularities in the past. This would mean that the Universe would have to start off in a smooth and uniform state of high order. However, the Weyl tensor would not, in general, be zero at singularities in the future which could be irregular and disordered.

There are several objections which can be raised to Penrose's proposal. First, it is rather *ad hoc*. Why should the Weyl tensor be zero on past singularities but not on future ones? In effect, one is putting in the thermodynamic arrow by hand. Second, it is based on the prediction of singularities in classical general relativity. However, it is generally believed that the gravitational field has to be quantized in order to be consistent with other field theories which are quantized. It is not clear whether singularities occur in quantum gravity or how to impose Penrose's boundary condition at them, if they do. Finally, Penrose's proposal does not explain why the cosmological and thermodynamic arrows should agree. With Penrose's boundary condition the thermodynamic arrow would agree with the cosmological arrow during the expanding phase of the Universe but it would disagree if the Universe were to start recollapsing.

The *CPT* invariance of the quantum state of the Universe defined by a path integral over compact metrics implies that if there is a certain probability of the Universe expanding, there must be an equal probability of it contracting. In order for the thermodynamic and cosmological arrows to agree in both the expanding and contracting phases, one requires boundary conditions which imply that the Universe is in a smooth state of high order when it is small but that it may be in an inhomogeneous disordered state when it is large. In Sec. IV it will be shown that the results of Ref. 11 imply that this is indeed the case for the quantum state defined by a path integral over compact metrics. This means that during the expansion phase the Universe starts out in a smooth state of high order but that, as it expands, it becomes more inhomogeneous and disordered. Thus the thermodynamic and cosmological arrows agree. However, when the Universe starts to recollapse, it has to get back to a smooth state when it is small. This means that disorder will decrease with time during the contracting phase and the thermodynamic arrow will be reversed. It will thus still agree with the cosmological arrow.

It should be emphasized that this reversal of the thermodynamic arrow of time is not caused by the gravitational fields or quantum effects at the point of maximum expansion of the Universe. Rather it is a result of the boundary condition that the Universe should be in a state of high order when it is small and it would occur in any theory which had this boundary condition as has been pointed out by a number of authors.[12,13] The only way that quantum gravity comes into the question of the arrow of time is that it provides a natural justification for the boundary condition.

One might ask what would happen to an observer (or computer) who survived from the expanding phase to the contracting one. One might think that one was free to enclose the observer or computer in a container that was so well insulated that he would be unaffected by the reversal of the thermodynamic arrow outside. If he were then to open a little window in his spaceship, he would see time going backward outside. The answer to this apparent paradox is that the observer's thermodynamic arrow, and hence his psychological arrow, would reverse at around the time of maximum expansion of the Universe, not because of effects that propagated into the spacecraft through the walls, but because of the boundary condition that the spacecraft be in a state of low entropy at late times when the Universe is small again. The contents of the memory of the observer or computer would increase during the expansion phase as the observer recorded observations but it would decrease during the contracting phase because the psychological arrow would be reversed and the observer would remember events in his future rather than his past.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

The prediction that the thermodynamic arrow would reverse if the Universe started to recontract may not have much practical importance because the Universe is not going to recollapse for a long time, if it ever does. However, we are fairly confident that localized regions of the Universe will collapse to form black holes. If one was in such a region, it would seem just like the whole Universe was collapsing around one. One might therefore expect that the region would become smooth and ordered, just like the whole Universe would if it recollapsed. Thus one would predict that the thermodynamic arrow of time should be reversed inside black holes. One would expect this reversal to occur only after one has fallen through the event horizon, so one would not be able to tell anyone outside about it. This and other consequences of the point of view adopted in this paper will be considered further in Sec. V. Section II will be a brief review of the canonical formulation of quantum gravity. In Sec. III it will be shown that the quantum state of the Universe defined by a path integral over compact metrics is invariant under *CPT*. Despite this invariance it will be shown in Sec. IV that the results of Ref. 11 imply that there is a thermodynamic arrow because the inhomogeneities in the Universe are small when the Universe is small but that they grow as the Universe expands.

II. CANONICAL QUANTUM GRAVITY

In the canonical approach the quantum state of the Universe is represented by a wave function $\Psi(h_{ij},\phi_0)$ which is a function of the three-metric h_{ij} and the matter field configuration ϕ_0 on a three-surface S. The interpretation of the wave function is that $|\Psi(h_{ij},\phi_0)|^2$ is the (unnormalized) probability of finding a three-surface S with three-metric h_{ij} and matter field configuration ϕ_0. The wave function is not an explicit function of time because there is no invariant definition of time in a curved space which is not asymptotically flat. In fact, the position in time of the surface S is determined implicitly by the three-metric h_{ij}. This means that $\Psi(h_{ij},\phi_0)$ obeys the zero-energy Schrödinger equation:

$$H\Psi(h_{ij},\phi_0)=0 . \tag{2.1}$$

This equation can be decomposed into two parts: the momentum constraint and the Wheeler-DeWitt equation. The momentum constraint is

$$\left[\frac{\delta\Psi}{\delta h_{ij}}\right]_{|i}=8\pi T^{0i}\Psi . \tag{2.2}$$

It implies that the wave function is the same on three-metrics h_{ij} and matter field configurations ϕ_0 that are related by a coordinate transformation. The Wheeler-DeWitt equation is

$$\left[-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}}+h^{1/2}(-{}^3R+16\pi T_{00})\right]\Psi=0 , \tag{2.3}$$

where

$$G_{ijkl}=\tfrac{1}{2}h^{-1/2}(h_{ik}h_{jl}+h_{il}h_{jk}-h_{ij}h_{kl}) .$$

It can be regarded as a second-order wave equation for Ψ on the infinite-dimensional space called superspace which is the space of all three-metrics h_{ij} and matter field configurations ϕ_0.

Any solution of Eqs. (2.2) and (2.3) represents a possible quantum state of the Universe. However, it seems reasonable to suppose that the Universe is not just in some arbitrary state but that its state is picked out or preferred in some way. As explained in Ref. 4, the most natural choice of quantum state is that for which the wave function is given by a path integral over compact metrics:

$$\Psi(h_{ij},\phi_0)=\int_C d[g_{\mu\nu}]d[\phi]\exp(-\tilde{I}[g_{\mu\nu},\phi]) , \tag{2.4}$$

where $\tilde{I}$ is the Euclidean action and the path integral is taken over four-metrics $g_{\mu\nu}$ and matter field configurations ϕ on compact four-manifolds which are bounded by the three-surface S with the induced three-metric h_{ij} and matter field configuration ϕ_0. The contour of integration in the space of all four-metrics has to be deformed from Euclidean (i.e., positive definite) metrics to complex metrics in order to make the path integral converge.[14,15] The proposal that the quantum state is given by (2.4) seems to give predictions that are in agreement with observation.[4,11,16]

III. THE *CPT* THEOREM

The precise statement of *CPT* invariance in flat spacetime is that the vacuum expectation values of bosonic quantum field operators $\phi(x)$ satisfy

$$\langle\phi(x_1)\phi(x_2)\cdots\phi(x_n)\rangle$$
$$=[\langle\phi^\dagger(-x_1)\phi^\dagger(-x_2)\cdots\phi^\dagger(-x_n)\rangle]^* . \tag{3.1}$$

In the case of fermion fields there is a factor of $(-1)^{F+J}$ where F is the fermion number and J is the number of undotted spinor indices. In the case of asymptotically flat spacetime one can formulate and prove *CPT* invariance in a similar way in terms of the vacuum expectation values of field operators at infinity.[17] However, although asymptotic flatness may be a reasonable approximation for local systems, one does not expect it to apply to the whole Universe. One therefore does not have any flat or asymptotically flat region in which one can define the *TP* operation $x\to -x$. All that one has is a wave function $\Psi(h_{ij},\phi_0)$ which is not an explicit function of time. However, one can introduce a concept of time by replacing the dependence of Ψ on $h^{1/2}$, the square root of the determinant of the three-metric h_{ij}, by its conjugate momentum, the trace of the second fundamental form of S. One defines the Laplace transform

$$\Phi(\tilde{h}_{ij},K_E,\phi_0)=\int_0^\infty d[h^{1/2}]\exp\left[-\frac{m_p{}^2K_E}{18\pi}\int K_E h^{1/2}d^3x\right]\Psi(h_{ij},\phi_0) , \tag{3.2}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

where $\tilde{h}_{ij}$ is the three-metric defined up to a conformal factor and K_E is the trace of the Euclidean second fundamental form. The Laplace transform Φ is holomorphic for $\mathrm{Re}(K_E)>0$. This means that one can analytically continue Φ in K_E to Lorentzian values $K_L=iK_E$ of the trace of the second fundamental form. Then $|\Phi(\tilde{h}_{ij},K_L,\phi_0)|^2$ is proportional to the probability of finding a three-surface S with the conformal three-metric $\tilde{h}_{ij}$, the rate of expansion K_L and the matter field configuration ϕ_0.

Consider first the case in which one has only fields like the gravitational field and real scalar fields which are invariant under C and P. The Euclidean action $\tilde{I}$ is real for Euclidean (i.e., positive definite) four-metrics $g_{\mu\nu}$ and real scalar fields ϕ. The contour of integration in the path integral (2.4) has to be deformed from Euclidean to complex metrics in order to make the integral converge. However, there will be an equal contribution from metrics with a complex action $\tilde{I}$ and from metrics with the complex conjugate action $(\tilde{I})^*$. Thus the wave function $\Psi(h_{ij},\phi_0)$ will be real. This implies that

$$\Phi(\tilde{h}_{ij},K_E,\phi_0)=\Phi^*(\tilde{h}_{ij},K_E^*,\phi_0) \tag{3.3}$$

for complex K_E. In particular, this implies

$$\Phi(\tilde{h}_{ij},K_L,\phi_0)=\Phi^*(\tilde{h}_{ij},-K_L,\phi_0) \tag{3.4}$$

for real K_L. Equation (3.4) is the statement of T invariance for the quantum state of the Universe. It implies that the probability of finding a contracting three-surface is the same as that of finding an expanding one, i.e., if the wave function represents an expanding phase of the Universe, then it will also represent a contracting one.

Consider now a situation in which one has charged fields, for example, a complex scalar field ϕ. The wave function Ψ will now be a functional of the three-metric h_{ij} and the complex field configuration ϕ_0 on S. In the Euclidean path integral (2.4) for Ψ one has to integrate over independent field configurations ϕ and $\tilde{\phi}$ on the Euclidean background $g_{\mu\nu}$ where $\phi=\phi_0$ and $\tilde{\phi}=\phi_0^*$ on S. The Euclidean action $\tilde{I}[g_{\mu\nu},\phi,\tilde{\phi}]$ is no longer necessarily real but

$$\tilde{I}[g_{\mu\nu},\phi,\tilde{\phi}]=\tilde{I}^*[g_{\mu\nu},\tilde{\phi},\phi]\ . \tag{3.5}$$

This implies

$$\Phi(\tilde{h}_{ij},K_L,\phi)=\Phi^*(\tilde{h}_{ij},-K_L,\phi^*)\ . \tag{3.6}$$

Equation (3.6) is a statement of the invariance of the quantum state of the universe under CT.

Finally one can consider fields, such as chiral fermions, which are not invariant under P. To deal with fermions one should introduce a triad of covectors e_i^a on S and should regard the wave function Ψ as a functional of the e_i^a and the fermion field ψ_0 on S. The path integral representation of the wave function is then

$$\Psi(e_i^a,\psi_0)=\int_C d[e_\mu^a]d[\psi]d[\bar{\psi}]\exp(-\tilde{I}[e_\mu^a,\psi,\bar{\psi}])\ , \tag{3.7}$$

where on S, $\psi=\psi_0$ and $\bar{\psi}=\bar{\psi}_0$. The oriented triad e_i^a on S defines a directed unit normal e_μ^0 to S. The path integral (3.7) is taken over all compact four-geometries which are bounded by S and for which e_μ^0 points inward.

The Euclidean action will obey

$$\tilde{I}[e_\mu^a,\psi,\bar{\psi}]=\tilde{I}^*[-e_\mu^a,\psi^C,\bar{\psi}^C]\ , \tag{3.8}$$

where $\psi^C=C\psi^*$ is the charge conjugate field and C is the charge conjugation matrix. This implies

$$\Psi(e_i^a,\psi)=\Psi^*(-e_i^a,\psi^C)\ . \tag{3.9}$$

One can regard (3.9) as the expression of the CPT invariance of the quantum state of the Universe because changing the sign of the triad e_i^a not only reverses the spatial directions, and so carries out the operation P, but it also reverses the direction of the orientated normal to S, e_μ^0. Alternatively, one can consider the Laplace transform Φ

$$\Phi(\tilde{e}_i^a,K_L,\psi)=\Phi^*(-\tilde{e}_i^a,-K_L,\psi^C)\ , \tag{3.10}$$

where $\tilde{e}_i^a$ is the triad in S defined up to a positive multiplicative factor.

It is clear that this proof of the CPT invariance of the quantum state defined by a path integral over compact metrics would apply equally well if there were higher derivative terms in the gravitational action. In the case of an action containing quadratic terms in the curvature, the wave function Ψ could be taken to be a function of the three-metric h_{ij}, the second fundamental form K^{ij}, and the matter field configuration ϕ_0. For fields that are invariant under C and P, the wave function $\Psi(h_{ij},K_E^{ij},\phi_0)$ would be real for real Euclidean values of the second fundamental form K_E^{ij}. This implies that

$$\Psi(h_{ij},K_L^{ij},\phi_0)=\Psi^*(h_{ij},-K_L^{ij},\phi_0)\ . \tag{3.11}$$

One can regard (3.11) as an expression of the T invariance of the quantum state. The extension to fields that are not invariant under C and P is straightforward. One can also apply similar arguments to the corresponding quantum state in Kaluza-Klein theories.

IV. THE INCREASE OF DISORDER

In Ref. 11 it was argued that the wave function $\Psi(h_{ij},\phi_0)$ can be approximated by a sum of terms of the form

$$\Psi_0(\alpha,\phi)\prod_n\Psi_n(\alpha,\phi,a_n,b_n,c_n,d_n,f_n)\ . \tag{4.1}$$

The wave function Ψ_0 describes a homogeneous isotropic closed Universe of radius e^α containing a homogeneous massive scalar field ϕ. The quantities $a_n,b_n,\ldots,f_n$ are the coefficients of harmonics of order n which describe perturbations from homogeneity and isotropy.

One can substitute (4.1) into the Wheeler-DeWitt equation and keep terms to all orders in the "background" quantities α and ϕ but only to second order in the "perturbations" $a_n,b_n,\ldots,f_n$. One obtains a second-order wave equation for Ψ_0 on the two-dimensional "minisuperspace" parametrized by the coordinates α and ϕ. The path integral (2.4) for the wave function implies that $\Psi_0\to 1$ as $\alpha\to-\infty$. One can integrate the wave equation with this boundary condition.[18] One finds that Ψ_0 starts to oscillate rapidly. This allows one to apply the WKB approximation

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$\Psi_0 = \mathrm{Re}(C e^{iS}) . \tag{4.2}$$

The trajectories of ∇S in the (α,ϕ) plane correspond to solutions of the classical field equations for a homogeneous isotropic Universe with a homogeneous massive scalar field. The trajectories corresponding to Ψ_0 start out at large values of $|\phi|$. They have a period of exponential expansion in which $|\phi|$ decreases followed by a period of matter dominated expansion in which ϕ oscillates around zero with decreasing amplitude. They reach a point of maximum expansion and then recontract in a time symmetric manner.

The perturbation wave functions Ψ_n can be further decomposed as follows:

$$\Psi_n = {}^S\Psi_n(\alpha,\phi,a_n,b_n,f_n)\,{}^V\Psi_n(\alpha,\phi,c_n)\,{}^T\Psi_n(\alpha,\phi,d_n) . \tag{4.3}$$

The wave function ${}^T\Psi_n$ describes gravitational wave perturbations parametrized by the coefficients d_n of the transverse traceless harmonics on the three-sphere. The wave function ${}^V\Psi_n$ describes the effect of gauge transformations which correspond to coordinate transformations on the three-sphere parametrized by the coefficients c_n of the vector harmonics. The wave function ${}^S\Psi_n$ parametrized by the coefficients a_n, b_n, and f_n of the scalar harmonics describe two gauge degrees of freedom and one physical degree of freedom of density perturbations. In situations in which the WKB approximation can be applied to the background wave function Ψ_0, the perturbation wave functions obey decoupled Schrödinger equations of the form

$$i\frac{\partial {}^T\Psi_n}{\partial t} = {}^T\!H\,{}^T\Psi_n , \tag{4.4}$$

where t is the time parameter of the solution of the classical field equations that corresponds to Ψ_0 via the WKB approximation.

One can evaluate the perturbation wave functions directly from the path integral expression (2.4) for the wave function. Consider, for example, the gravitational wave perturbations. One can regard them as quantum fields parametrized by d_n' propagating on a homogeneous isotropic background metric of the form

$$ds^2 = -N(t)^2dt^2 + e^{2\alpha(t)}d\Omega_3{}^2 , \tag{4.5}$$

where $d\Omega_3{}^2$ is the metric on the unit three-sphere, if the lapse function N is real everywhere, the metric (4.5) has a Lorentzian signature and cannot be compact and nonsingular. However, I shall consider complex background fields $(N(t),\alpha'(t),\phi'(t))$ such that at some value $t=t_0$, N is negative imaginary. The metric then has a Euclidean signature at $t=t_0$ and will be regular and compact if $\alpha' = -\infty$, $d\alpha'/dt = iN e^{-\alpha'}$, and $d_n' = 0$. The argument of N will vary continuously with t. When N becomes real, the metric will become Lorentzian. One can express the perturbation wave functions as path integrals on these backgrounds, e.g.,

$${}^T\Psi_n(\alpha,\phi,d_n) = \int d[d_n']\exp(-\bar{I}[\alpha',\phi',d_n']) , \tag{4.6}$$

where the path integral is taken over all gravitational wave perturbations d_n' on all regular compact background fields described by $\alpha'(t)$ and $\phi'(t)$.

The path integral over d_n' in a given background field is Gaussian and therefore can be evaluated as

$$(\det\Delta)^{-1/2}\exp(-\bar{I}^{\mathrm{cl}}[d_n]) , \tag{4.7}$$

where Δ is a differential operator and

$$\frac{e^{3\alpha}}{2iN}\left[d_n'\frac{d}{dt}d_n' + 4\frac{d\alpha'}{dt}d_n'^2\right] \tag{4.8}$$

is the action of a solution of the classical field equations for a perturbation d_n' on the given background with $d_n' = 0$ at $t=t_0$ and $d_n' = d_n$ at the location $t=t_1$ of the three-surface S.

One expects the dominant contribution to the path integral (4.6) to come from backgrounds which are close to solutions of the classical background equations. These solutions will be Euclidean (N imaginary) at $t=t_0$ and they will become Lorentzian in those regions of the (α,ϕ) plane in which Ψ_0 oscillates and the WKB approximation can be applied. In such a background the classical field equation for d_n' is

$$\left[-\frac{d}{dt}\left[\frac{e^{3\alpha'}}{iN}\frac{d}{dt}\right] + iN e^{\alpha'}(n^2-1)\right]d_n' = 0 . \tag{4.9}$$

In the region of the (α',ϕ') plane in which the WKB approximation can be applied and N is real, one can regard Eq. (4.9) as a harmonic oscillator equation for the variable $x = \exp(3/2\alpha')d_n'$ with the time-dependent frequency $\nu = \exp(-\alpha')(n^2-1)^{1/2}$. If α' were independent of t, the solution of (4.9) that obeys the above boundary conditions is

$$d_n' = d_n\frac{\sin\nu\tau}{\sin\nu\tau_1} , \tag{4.10}$$

where $\tau = \int_{t_0}^{t} N\,dt$.

Of course α' will vary with t but (4.10) will still be a good approximation provided that the adiabatic approximation holds, i.e., $|\dot{\alpha}'/N|$, the rate of change of α', is small compared to the frequency ν. In the Euclidean region near $t=t_0$, this will be true because $|\dot{\alpha}'/N| < e^{-\alpha'}$. In the Lorentzian region it will be true for perturbation modes whose wavelength ν^{-1} is small compared to the horizon distance $N/\dot{\alpha}'$. For such modes

$$\frac{d}{dt}d_n' = N\nu d_n'\cot\nu\tau . \tag{4.11}$$

For t_1 in the region in which the WKB approximation can be applied and for $n \gg 1$, the imaginary part of $\nu\tau_1$, which arises from the Euclidean region near $t=t_0$, will be less than $-i$. This means that the real part of the Euclidean action (4.8) will be $\frac{1}{2}\nu e^{3\alpha}d_n{}^2 = \frac{1}{2}\nu x^2$. The imaginary part of the Euclidean action will be small. It will give rise to a phase factor in ${}^T\Psi_n$ which can be removed by a canonical transformation of variables. Thus the perturbation wave function will have the ground-state form

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$^{T}\Psi_n(d_n) = B \exp[-\tfrac{1}{2}\nu \exp(3\alpha) d_n^{\,2}]$$

$$= B\, e^{-\nu x^2/2} . \qquad (4.12)$$

The vector perturbation wave function $^{V}\Psi_n(c_n)$ describes a gauge degree of freedom and does not have any physical significance. The scalar perturbation, which is a function $^{S}\Psi_n(a_n, b_n, f_n)$ describes two gauge degrees of freedom and one physical degree of freedom. A similar analysis and use of the adiabatic approximation shows that this physical degree of freedom is in its ground state when the wavelength of the perturbation is less than the horizon size during the period of exponential expansion. Thus at early times in the exponential expansion, i.e., when the Universe is small, the physical perturbation modes of the Universe have their minimum excitation. The Universe is in a state that is as ordered and homogeneous as it can be consistent with the uncertainty principle. This ordered state is not only an initial state for the expansion phase of the Universe but it is also a final state for the contracting phase because the WKB trajectories for Ψ_0 return to the same region of the (α,ϕ) plane and the perturbation wave functions depend only on the position in this plane.

On the other hand, the perturbation modes are not in their ground state when the Universe is large because in this case the adiabatic approximation breaks down when the wavelength of the perturbation becomes greater than the horizon size during the period of exponential expansion. Detailed calculations[11] show that when the scalar perturbation modes reenter the horizon during the matter-dominated era, they are in a highly excited state and give rise to a scale-free spectrum of density fluctuations $\delta\rho/\rho$. These density inhomogeneities provide the initial conditions necessary for the formation of galaxies and other structures in the Universe. The perturbation wave functions are still in a very special state because their phase factors have to be such that when they are evolved according to the Schrödinger equation, they will return to their ground-state form when the Universe recontracts. However, this special nature of the perturbation wave functions would not be noticed by an observer who makes the usual coarse-grained measurements. All he would notice was that during the expansion the Universe had evolved from a homogeneous, ordered state to an inhomogeneous, disordered state. Thus he would say that the thermodynamic arrow pointed in the direction of time in which the Universe was expanding. On the other hand, an observer in the contracting phase would feel that the Universe was evolving from a state of disorder to one of order. He would therefore ascribe the opposite direction to the thermodynamic arrow and would also find that it agreed with the cosmological arrow.

The connection between the thermodynamic and cosmological arrows should hold in models that are more general than the one considered in Ref. 11 because it depends only on the fact that the adiabatic approximation should hold for small perturbations on "small" three-geometries but not for perturbations on "large" three-geometries. Thus one might expect that it would also hold in models that allowed for the formation of black holes as a result of the gravitational collapse of density fluctuations produced during the expansion. This would mean that the thermodynamic arrow would reverse inside a black hole. This is currently under investigation.

V. CONSEQUENCES

Are there any observable consequences of the prediction that the thermodynamic arrow should reverse in a recontracting phase of the Universe or inside a black hole? Of course, one could wait until the Universe recollapsed or one could jump into a black hole. However, the probability distribution of the density parameter $\Omega = \rho/\rho_{\text{crit}}$ seems to be concentrated at $\Omega = 1$ (Ref. 16). Thus one would have to wait a very long time for the collapse of the Universe. On the other hand, if one jumped into a black hole, one would not be able to tell anyone outside. Furthermore, if the thermodynamic arrow did reverse, one would not remember it because it would now be in one's future rather than in the past.

In principle it is possible to determine from the present positions and velocities of clusters of galaxies that they developed from an initial configuration with very low peculiar velocities. In a similar way it should therefore be possible to calculate whether they will evolve to a state with low peculiar velocities at some time in the future. The difficulty is that on the basis of the inflationary model, one would expect the value of Ω for the presently observed Universe to be equal to one to one part in 10^4. Thus one would expect the Universe to expand by a further factor of at least 10^4 before it began to recontract. In this extra expansion other clusters of galaxies which we have not yet observed would appear over the horizon and their gravitational fields could have a significant effect on the behavior of clusters near us. Thus it would seem very difficult to make an experimental test of the prediction that the thermodynamic arrow would reverse if the Universe began to recontract.

A better bet would seem to be to study the inflow of matter into a black hole. At least in principle this is a situation that we ought to be able to observe with some accuracy. However, on the basis of classical general relativity, one might expect the boundary of the region of high spacetime curvature not to be spacelike, as it is in the Schwarzschild solution, but to be null, like the Cauchy horizon in the Reissner-Nordström or Kerr solutions. If this were the case, the behavior of the matter and metric on the brink of the quantum era would depend on the entire future history of infall into the black hole. Merely to observe the infall for a limited period of time would be insufficient to determine whether or not the thermodynamic arrow of time reversed near the region of high curvature. Clearly more work has to be done on the classical and quantum aspects of gravitational collapse.

One might think that the *CPT* theorem implied that all the baryons in the Universe would have to decay into leptons before the Universe began to recollapse and that the leptons would be reassembled into antibaryons in the collapsing phase. If this were the case, one could disprove the proposed "no boundary" condition for the Universe if one could show that the observed value of Ω was such that the Universe should begin to recollapse before all the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

baryons had decayed. However, what the *CPT* theorem implies is just that the probability of finding an expanding three-surface with a matter configuration of baryons is the same as that of finding a contracting three-surface with a matter configuration of antibaryons. This requirement is no restriction at all because the two three-surfaces can merely be the same three-surface viewed with different orientation of time: reversing the orientation of time and space interchanges the labels, baryons, and antibaryons. Thus the *CPT* invariance of the quantum state of the Universe does not imply any limit on the lifetime of the proton. In any case, we certainly do not observe baryons changing into antibaryons as they fall into a black hole.

To sum up, the proposal that spacetime is compact without boundary implies that the quantum state of the Universe is invariant under *CPT*. Despite this, one would observe an increase in (coarse-grained) entropy during an expansion phase of the Universe. However, it seems difficult to test the prediction that entropy should decrease during a contracting phase of the Universe or inside a black hole.

Note added in proof. Since this paper was submitted for publication a paper by Don Page has appeared [following paper, Phys. Rev. D 32, 2496 (1985)]. In it he questions my conclusion that the thermodynamic arrow of time would reverse in a contracting phase of the universe or in a black hole. My conclusion was based on the fact that the wave function Ψ went exactly into 1 as one goes to $\alpha = -\infty$ on a null geodesic in the α,ϕ plane. This would imply that Ψ was not oscillating at large negative α and therefore that all the classical Lorentzian contracting solutions would have to bounce at a small radius. At the bounce one could apply an analysis similar to that in Ref. 11 to show that all the inhomogeneous modes were in their ground state. This would mean that the inhomogeneity would decrease in the collapsing phase and therefore that the thermodynamic arrow of time would be reversed.

Page has pointed out however that even at large negative α, there might be a small oscillating component in the wave function. This would arise from complex stationary points in the path integral over compact metrics that were near to the Lorentzian metric which started with an inflationary expansion, reached a maximum radius and then recollapsed to zero radius without boundary. Although the amplitude of this oscillating component would be small, its frequency would be very high. It would therefore correspond to an appreciable probability flux of classical solutions in the WKB approximation.[16] One would not expect the inhomogeneous perturbations about such solutions to be in their ground state when the solution recollapsed because the adiabatic approximation used in Ref. 11 would break down. There is thus no reason for the thermodynamic arrow of time to reverse in these solutions. Similarly one would not expect it to reverse inside black holes.

I think that Page may well be right in his suggestion. In that case the two main results of this paper that are correct are, first, that the wave function is invariant under *CPT*, though this does not imply that the individual classical solutions that correspond to the wave function via the WKB approximation are invariant under *CPT*, second, that the classical solutions, which start out with an inflationary period, will have a well-defined thermodynamic arrow of time.

---

[1]R. F. Streater and A. S. Wightman, *PCT, Spin, Statistics and All That* (Benjamin, New York, 1964).

[2]S. W. Hawking, in *Astrophysical Cosmology: Proceedings of the Study Week on Cosmology and Fundamental Physics*, edited by H. A. Brück, G. V. Coyne, and M. S. Longair (Pontificiae Academiae Scientiarum Scripta Varia, Vatican City, 1982), pp. 563–574.

[3]J. B. Hartle and S. W. Hawking, Phys. Rev. D 28, 2960 (1983).

[4]S. W. Hawking, Nucl. Phys. B239, 257 (1984).

[5]J. E. Hogarth, Proc. R. Soc. London A267, 365 (1962).

[6]F. Hoyle and J. V. Narlikar, Proc. R. Soc. London A277, 1 (1964).

[7]J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. 17, 157 (1945); 21, 425 (1949).

[8]*The Nature of Time*, edited by T. Gold and D. L. Schumacher (Cornell University Press, Ithaca, 1967).

[9]R. Penrose, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, England, 1979).

[10]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge University Press, England, 1973).

[11]J. J. Halliwell and S. W. Hawking, Phys. Rev. D 31, 1777 (1985).

[12]P. C. W. Davies, *The Physics of Time Asymmetry* (Surrey University Press/California University Press, Berkeley, 1974), Sec. 7.4.

[13]T. Gold, in *La Structure et l'Evolution de l'Universe*, 11th International Solvay Congress (Edition Stoops, Brussels, 1958); Am. J. Phys. 30, 403 (1962); in *Recent Developments in General Relativity* (Pergamon-MacMillan, New York, 1962); H. Bondi, Observatory 82, 133 (1962); D. L. Schumacher, Proc. Cambridge Philos. Soc. 60, 575 (1964); M. Gell-Mann, comments in Proceedings of the Temple University Panel on Elementary Particles and Relativistic Astrophysics (unpublished); Y. Ne'eman, Int. J. Theor. Phys. 3, 1 (1970); P. T. Landsberg, Stud. Gen. 23, 1108 (1970); W. J. Cocke, Phys. Rev. 160, 1165 (1967); H. Schmidt, J. Math. Phys. 7, 495 (1966).

[14]G. W. Gibbons, S. W. Hawking, and M. J. Perry, Nucl. Phys. B138, 141 (1978).

[15]S. W. Hawking, in *Relativity, Groups and Topology*, Les Houches, 1983, edited by B. S. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).

[16]S. W. Hawking and D. N. Page, report 1985 (unpublished).

[17]S. W. Hawking, Commun. Math. Phys. 87, 395 (1982).

[18]S. W. Hawking and Z. C. Wu, Phys. Lett. 151B, 15 (1985).

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

Nuclear Physics B264 (1986) 185–196
© North-Holland Publishing Company

OPERATOR ORDERING AND THE FLATNESS OF THE UNIVERSE

S.W. HAWKING

University of Cambridge, Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, UK

D.N. PAGE[1]

Physics Department, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Received 23 May 1985

In order to make predictions in quantum cosmology one has to resolve the problem of the operator ordering in the hamiltonian. This is equivalent to defining a differential operator on superspace, the infinite dimensional manifold of all 3-metrics and matter field configurations on a 3-surface S. We propose that this operator should be the laplacian in a natural metric. There remains a difficulty however because the metric connection on superspace depends on the metric in a nonlinear manner. This nonlinearity would be inconsistent with the interpretation of the lapse function as a Lagrange multiplier. It may cancel out if an equal number of fermion degrees of freedom are included. The probability measure on superspace would be the measure associated with this metric. In situations in which the wave function can be interpreted in terms of classical solutions by the WKB approximation, this choice of measure implies that the probability of finding the 3-metric and matter field configuration in a given region of superspace is proportional to the proper time that the solutions spend in that region. As an application we compute the probability distribution of the density parameter Ω in a minisuperspace model which is in the quantum state defined by a path integral over compact 4-geometries. If we restrict attention to a fixed value of the density, we find that the probability distribution is entirely concentrated at $\Omega = 1$.

1. Introduction

Observations indicate that the large-scale structure of the Universe at the present time is very close to that of a Friedmann-Robertson-Walker model with nearly or exactly the critical density. In an attempt to account for this Starobinsky [1] and Guth [2] proposed models in which the Universe would have undergone a period of exponential or "inflationary" expansion in the early stages driven by quantum corrections or by the vacuum energy of the unbroken symmetry phase of a grand unified theory (see [3]). The exponential expansion would reduce any anisotropy or inhomogeneity which was present and would make the density parameter $\Omega = \rho/\rho_{\text{crit}}$ very close to one. However, this model fails as a complete explanation of the present state of the Universe because it does not make any statement about the initial boundary conditions of the Universe. It indicates that there could be many sets of initial conditions which lead to something like the present state but it cannot show that *every* set leads to the present state because one could pick an arbitrary

[1] Member at The Institute for Advanced Study, Spring 1985.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

inhomogeneous anisotropic state for the Universe at the present time and evolve it backwards in time using the field equations to find out what initial conditions it would have arisen from. In the absence of a probability measure on the space of all initial conditions, one could not even say that the present state of the Universe was probable, let alone that it would be necessary.

In order to account for the present state of the Universe, one has to make some assumption about the boundary conditions of the Universe. It has recently been proposed [4-7] that the boundary condition of the Universe is that it has no boundary. In other words, the quantum state of the Universe is defined by a path integral over compact metrics without boundary. In simple minisuperspace examples [7-10] in which the number of degrees of freedom of the gravitational and matter fields are restricted to a finite number it was shown that this boundary condition led to a quantum state that could be regarded as a superposition of quantum states that were peaked around a family of solutions of the classical field equations that had a period of exponential expansion. However in order to use these models to make predictions about the probability of the Universe being as we observe it, one has to resolve the related issues of the operator ordering in the Wheeler-DeWitt equation for the wave function of the Universe and the probability measure on superspace, the space of all 3-metrics and matter field configurations. The aim of this paper is to propose a natural solution to these questions. In situations in which the wave function can be interpreted in terms of classical solutions by the WKB approximation, this prescription implies that the probability of finding a three geometry within certain limits is proportional to the lengths of proper time that the solutions spend between the limits. As an application, we show that these minisuperspace models predict a probability distribution for the density parameter Ω at a given density that is entirely concentrated at $\Omega = 1$. Thus the "no boundary" proposal for the quantum state of the Universe seems to make the testable (and falsifiable) prediction that the Universe should have exactly the critical density.

In sect. 2 we present our solution to the operator ordering problem for the Wheeler-DeWitt equation and point out a difficulty with it that might be resolved by the inclusion of fermions. Sect. 3 contains a discussion of the probability interpretation of the wave function of the Universe. In sect. 4 we apply the results of the two previous sections to minisuperspace models.

2. The Wheeler–DeWitt equation

The metric of spacetime can be written locally in the 3+1 form

$$ds^2 = -(N^2 - N_i N^i)\,dt^2 + 2N_i\,dt\,dx^i + h_{ij}\,dx^i\,dx^j, \tag{2.1}$$

where x^i ($i = 1, 2, 3$) are coordinates on a three-dimensional manifold S and indices i and j are raised and lowered with the three-metric h_{ij}. The hamiltonian of classical

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

general relativity is

$$H = \int (NH_0 + N_i H^i)\, \mathrm{d}^3x, \tag{2.2}$$

where

$$H_0(x) = \int \Big[16\pi m_{\mathrm{p}}^{-2} G_{ijkl}(x, x')\pi^{ij}(x)\pi^{kl}(x') + h^{1/2}\delta^3(x-x')\left(\frac{m_{\mathrm{p}}^2}{16\pi}(-{}^3R + 2\Lambda) + T_{00}(\phi, \pi_\phi, h_{ij})\right)\Big] \mathrm{d}^3x, \tag{2.3}$$

$$G_{ijkl}(x, x') = \tfrac{1}{2}h^{-1/2}[h_{ik}(x)h_{jl}(x') + h_{il}(x)h_{jk}(x') - h_{ij}(x)h_{kl}(x')]\delta^3(x-x'), \tag{2.4}$$

$$H^i(x) = 2\pi^{ij}{}_{|j} - h^{1/2}T^{0i}, \tag{2.5}$$

and

$$\pi^{ij} = -\frac{m_{\mathrm{p}}^2}{16\pi} h^{1/2}(K^{ij} - h^{ij}K) \tag{2.6}$$

is the momentum conjugate to h_{ij}. The 0 indices on the energy-momentum operator T in (2.3) and (2.5) refer to an orthonormal basis but the i index in (2.5) refers to a coordinate basis.

The lapse N and shift N_i are regarded as Lagrange multipliers. Thus the classical field equations are

$$H_0(x) = 0, \qquad H^i(x) = 0 \tag{2.7}$$

for all x, or equivalently,

$$H = 0 \tag{2.8}$$

for all N and N_i.

The quantum state of a gravitational system can be described by a wave function Ψ which is a functional of the three-metric h_{ij} and the matter field configuration ϕ on S. For simplicity we shall confine attention to the case of a scalar field ϕ but the extension to other matter fields is straightforward. In this case Ψ is a function on the infinite dimensional manifold W called superspace which consists of all positive-definite three-metrics $h_{ij}(x)$ and the matter fields $\phi(x)$ on S. A vector in the tangent space at a point of W consists of a pair $(\gamma_{ij}(x), \mu(x))$ where $\gamma_{ij}(x)$ is a symmetric tensor field and $\mu(x)$ is a scalar field on S. One can regard γ_{ij} as a perturbation on

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

h_{ij} and μ as a perturbation on ϕ. For each $N>0$ on S there is a natural metric $I(N)$ on W defined by

$$ds^2 = \int\int N(x)^{-1}\left[\frac{m_p^2}{32\pi} G^{ijkl}(x,x')\gamma_{ij}(x)\gamma_{kl}(x') + h^{1/2}\delta^3(x-x')\mu(x)\mu(x')\right] d^3x\, d^3x' \tag{2.9}$$

where

$$G^{ijkl}(x,x') = \tfrac{1}{2}h^{1/2}[h^{ik}(x)h^{jl}(x') + h^{il}(x)h^{jk}(x') - 2h^{ij}(x)h^{kl}(x')]\delta^3(x-x') \tag{2.10}$$

is the inverse of $G_{ijkl}(x,x')$.

The wave function obeys the zero-energy Schrödinger equation

$$H\Psi = 0, \tag{2.11}$$

where the momenta π^{ij} and π_ϕ have been replaced by $-i\delta/\delta h_{ij}$ and $-i\delta/\delta\phi$ respectively. Because the lapse N and shift N_i are independent Lagrange multipliers, one can split this into the Wheeler-DeWitt equation

$$H_| \Psi = 0, \tag{2.12}$$

where

$$H_| = \int N H_0\, d^3x \tag{2.13}$$

and the momentum constraints

$$H_- \Psi = 0, \tag{2.14}$$

where

$$H_- = \int N_i H^i\, d^3x. \tag{2.15}$$

The momentum constraint implies that the gradient of Ψ on W is zero in the directions that correspond to gauge transformations i.e. diffeomorphisms of S.

There is one Wheeler-DeWitt equation for each field N on S. Each is a second order differential equation for Ψ on W. There is some ambiguity in the ordering when the momenta are replaced by differential operators so in general one has

$$H_| = -\tfrac{1}{2}\nabla^2 + \beta\int m_p^{-2} N h^{-1/2}\, d^3x\, \gamma_{ij}\frac{\delta}{\delta h_{ij}} + \varepsilon + V, \tag{2.16}$$

where

$$V = \int N h^{1/2}\left[\frac{m_p^2}{16\pi}(-{}^3R + 2\Lambda) + U(\phi)\right] d^3x. \tag{2.17}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

∇^2 is the laplacian in the metric $\Gamma(N)$, $(\gamma_{ij}, 0)$ is a vector field on W, ε is a scalar on W and $U = T_{00} - \frac{1}{2}\pi_\phi^2$. The only natural choice for ε is $\varepsilon = \xi\mathbb{R} + \eta \int Nh^{1/2}\, d^3x$ where ξ and η are constants and $\mathbb{R}$ is the scalar curvature of W in the metric $\Gamma(N)$. One can absorb η into a renormalization of the cosmological constant Λ.

The only natural choice for the vector field $(\gamma_{ij}, 0)$ is $(h_{ij}, 0)$. However, if the coefficient β of the first derivative term were nonzero, the operator $H_|$ would not be a locally self-adjoint differential operator in the natural measure generated by the metric $\Gamma(N)$. It is therefore reasonable to put $\beta = 0$. Rather similar suggestions have been made in [11] and [12]. For the omission of first derivative terms to be consistent for all the Wheeler-DeWitt equations corresponding to different N one has to regard N as a field on S which is independent of h_{ij}. This means that one has to use a hamiltonian which is conjugate to some (possibly spatially variable) multiple of proper time rather than a hamiltonian that was conjugate say to conformal time for which N would be proportional to $h^{1/6}$. The most natural choice of the coefficient ξ of the scalar curvature $\mathbb{R}$ is zero but we shall bear in mind the possibility of other values.

The prescription for the operator ordering described above has the virtue that it defines a Wheeler-DeWitt operator which is independent of the choice of coordinates on superspace. However, it suffers from the problem that the connection on superspace defined by the metric $\Gamma(N)$ does not depend linearly on N. This means that the hamiltonian operator is not a linear function of N as it would have to be in order that N can be interpreted as a Lagrange multiplier. It may be that the nonlinear dependence on N would cancel out in theories like supergravity which contain equal numbers of fermionic and bosonic degrees of freedom. This possibility is suggested by the work of Fradkin and Vilkovisky [13] who found that the measure in the path integral seemed to contain a non-covariant dependence on g^{00} (which equals $-N^{-2}$ in our notation) However, this term cancelled out in theories with equal numbers of bosons and fermions.

3. Probability interpretation

There has been some confusion about the probability interpretation of the wave function Ψ. If one chooses $\beta = 0$, so that the operator $H_|$ is self-adjoint, the current

$$J = \tfrac{1}{2}i(\Psi\nabla\bar{\Psi} - \bar{\Psi}\nabla\Psi) \tag{3.1}$$

is conserved in each of the metrics $\Gamma(N)$. However, one cannot interpret J directly as a probability flux because it is not positive definite. Indeed J vanishes for the quantum state defined by a path integral over compact metrics because Ψ is real in this case. Instead we shall adopt the approach, described in [7], that $|\Psi|^2$ is proportional to the probability of finding a 3-surface S with metric h_{ij} and matter

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

field configuration ϕ. In other words, the probability of finding a 3-metric and matter field configuration in a region A of W is

$$P(\mathrm{A}) \propto \int_{\mathrm{A}} |\Psi|^{2} * 1 , \tag{3.2}$$

where * is the Hodge dual in the metric $\Gamma(N)$, i.e. *1 is the volume element. One might think that there was ambiguity in the probability measure on W because of all the different metrics $\Gamma(N)$ but for N independent of h_{ij} and for Ψ which satisfy the momentum constraint, all these probability measures are equivalent in the sense that the right-hand sides of (3.2) differ only by a constant factor independent of position in W. This is another reason to choose N to be independent of h_{ij}.

In certain situations Ψ can be interpreted in terms of classical solutions by the WKB approximation. One puts

$$\Psi = \Psi_{+} + \Psi_{-} , \tag{3.3}$$

$$\Psi_{+} = \bar{\Psi}_{-} = \sum_{n} \Psi_{n} = \sum_{n} C_{n} \, \mathrm{e}^{iS_{n}} , \tag{3.4}$$

where the phase factors S_n vary much more rapidly on W than the prefactors C_n. One chooses the phase factors S_n to satisfy the classical Hamilton-Jacobi equation:

$$\tfrac{1}{2}(\nabla S_{n})^{2} + V = 0 , \tag{3.5}$$

where $(\nabla S_n)^2$ is the square of the gradient of S_n in the metric $\Gamma(N)$. Then the Wheeler-DeWitt equation for Ψ_n will be satisfied if

$$\tfrac{1}{2}\nabla^{2} C_{n} + i(\nabla S_{n}) \cdot (\nabla C_{n}) + \tfrac{1}{2} i C_{n} \nabla^{2} S_{n} = 0 . \tag{3.6}$$

The WKB approximation is to neglect the first term in (3.6) in comparison with the terms involving derivatives of S_n. One can then solve equation (3.6) for C_n along the integral curves of X_n where $X_n = \partial/\partial t$ is the vector field on W obtained by raising the covector ∇S_n by the inverse to the metric $\Gamma(N)$ and t is the time coordinate of the classical solution of the field equations corresponding to the Hamilton-Jacobi function S_n and the choice of N.

Each WKB component Ψ_n will have a conserved current

$$J_{n} \approx |C_{n}|^{2} \nabla S_{n} , \tag{3.7}$$

which flows along the directions of the classical trajectories. Consider a thin pencil B of the trajectories. The flux across a hypersurface Σ of codimension one in W

$$F_{n}(\mathrm{B}) = \int_{\Sigma \cap \mathrm{B}} * J_{n} \tag{3.8}$$

will be independent of Σ. The contribution of the pencil B to the probability of the

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

surface S lying within the region A of superspace will be

$$P_n(\mathrm{A}\cap\mathrm{B}) \propto \int_{\mathrm{A}\cap\mathrm{B}} |\Psi_n|^{2*}1 = \int_{\mathrm{A}\cap\mathrm{B}} |C_n|^{2*}1$$

$$= \int_{\mathrm{A}\cap\mathrm{B}} \mathrm{d}t \wedge {}^*J_n$$

$$= F_n(\mathrm{B}) \int \mathrm{d}t. \tag{3.9}$$

Thus the contribution of the pencil B to the probability $P(\mathrm{A})$ is proportional to the coordinate time Δt that the classical solutions in B spend in the region A. As was shown in the previous section, the requirement that $H_|$ be self-adjoint implies that N has to be independent of h_{ij}. Thus Δt is proportional to proper time. In other words, the probability of finding a 3-metric h_{ij} and matter field configuration ϕ lying in a region A of superspace is proportional to the proper times that are spent in A by the classical solutions that correspond to the wave function in the WKB approximation. This result would not have been true if we had chosen β not to be zero.

4. Application to minisuperspace

The full wave function for all the dynamical variables in the Universe would be very difficult to calculate. Attention, therefore, has been concentrated on minisuperspace models in which the infinite number of degrees of freedom of the gravitational and matter fields are restricted to a finite number. In this section we shall examine the consequences for some of these models of our proposal for the operator ordering and probability measure. We shall calculate the conditional probability distribution for the density parmeter Ω at a given density and shall show that it is entirely concentrated at $\Omega = 1$. Thus these models make the testable prediction that Ω should be one.

We consider models in which the gravitational field is restricted to be locally homogeneous and isotropic. Thus once a suitable topology of the compact 3-surface S is chosen, the 3-metric is determined by a single scale length a and may be written as

$$h_{ij} = \sigma^2 a^2 \tilde{h}_{ij}, \tag{4.1}$$

where the unit 3-space metric $\tilde{h}_{ij}$ has curvature

$${}^3\tilde{R}_{ijkl} = k(\tilde{h}_{ik}\tilde{h}_{jl} - \tilde{h}_{il}\tilde{h}_{jk}) \tag{4.2}$$

with $k = +1$ (for S a 3-sphere or a 3-sphere factored by a discrete group), $k = 0$ (for a 3-torus or other flat space), or $k = -1$ (for a hyperbolic space factored by a discrete

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

group). For convenience we have chosen the normalization constant σ to be

$$\sigma = \left(\frac{3m_p^2}{4\pi} \int \tilde{h}^{1/2} \, d^3x \right)^{-1/2} . \tag{4.3}$$

The matter content will be taken to be a single scalar field ϕ which is constant on the surfaces of homogeneity. In order to simplify the equations it will be expressed in the form

$$\phi = \left(\frac{3m_p^2}{4\pi} \right)^{1/2} \tilde{\phi} . \tag{4.4}$$

The self-interaction potential $U(\phi)$ will be written as

$$U(\phi) = \frac{3m_p^2}{4\pi\sigma^2} \tilde{U}(\tilde{\phi}) . \tag{4.5}$$

For example, if ϕ is a free field with mass $\mu = \sigma^{-1}m$, then $\tilde{U} = \frac{1}{2}m^2\tilde{\phi}^2$, but other forms of $\tilde{U}(\tilde{\phi})$ will be considered. We shall assume that the cosmological constant is zero.

The minisuperspace $\tilde{W}$ is two-dimensional with coordinates a and $\tilde{\phi}$. For an N which is constant on the surfaces of homogeneity, the minisuperspace metric $\Gamma(N)$ will be

$$ds^2 = \sigma N^{-1}(-a \, da^2 + a^3 \, d\tilde{\phi}^2) . \tag{4.6}$$

The Wheeler–DeWitt equation is

$$\left(a\frac{\partial}{\partial a} a\frac{\partial}{\partial a} - \frac{\partial^2}{\partial \tilde{\phi}^2} - ka^4 + 2a^6\tilde{U} \right) \Psi = 0 . \tag{4.7}$$

In this two-dimensional minisuperspace model, there is no ambiguity from the $\xi\mathbb{R}$ term, because the minisuperspace metric (4.6) is flat for N independent of $\tilde{\phi}$. As we have argued above, N should also be independent of h_{ij}, which in this homogeneous case restricts it to be a constant that for simplicity will be chosen to be σ. Then $\Gamma(N = \sigma)$ is the metric for the interior of the future light cone of the origin in a two-dimensional Minkowski spacetime with null coordinates

$$u = T - Z = \tfrac{2}{3}a^{3/2} e^{-3\tilde{\phi}/2} , \tag{4.8}$$

$$v = T + Z = \tfrac{2}{3}a^{3/2} e^{+3\tilde{\phi}/2} . \tag{4.9}$$

The Wheeler–DeWitt equation (4.7) is hyperbolic. Thus one can determine its solution in the interior of the future light cone, $T \geq |Z|$, if one knows the value of Ψ on the future light cone $T = |Z|$. This can in principle be determined for the quantum state defined by a path integral over compact 4-geometries without boundary. Although it is difficult to evaluate the path integral directly, one would expect it to give a wave function which is nearly constant for small uv if the corresponding 3-space configuration can be filled in with a compact 4-space with negligible

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

gravitational and matter actions. For a sequence of such 3- and 4-spaces of constant shape but linear size proportional to a, the gravitational action scales as a^2, so it should be negligible for sufficiently small a. The matter action would also go to zero with a if one considered a sequence of 4-spaces with the same distribution of matter fields within each, so the wave function should tend to a constant for fixed $\tilde{\phi}$ as a goes to zero. However, this gives the wave function only near $u = v = 0$. To obtain Ψ along the entire light cone, one needs to consider a sequence of 4-space configurations, bounding 3-space configurations $(a, \tilde{\phi})$ in which

$$T = \tfrac{1}{2}(u+v) = \tfrac{2}{3}a^{3/2}\cosh(\tfrac{3}{2}\tilde{\phi}) \tag{4.10}$$

is kept constant along the sequence as $uv = \frac{9}{4}a^3$ is reduced to zero. One such sequence in which the 4-spaces all have the same shape is given by the euclidean 4-metrics

$$ds^2 = \sigma^2 a^2[d\tau^2 + \tau^2 \tilde{h}_{ij}\, dx^i\, dx^j], \qquad 0 \leqslant \tau \leqslant 1. \tag{4.11}$$

At $\tau = 1$ the boundary is the three metric (4.1). If this is a 3-sphere, the 4-metric (4.11) is simply flat 4-space with $\tau = 0$ being the centre, a regular point. If $\tilde{h}_{ij}$ is not a 3-sphere, (4.11) will be singular at $\tau = 0$ but the topology and 4-metric may be altered in a small region there to remove the singularity. The total gravitational action after this is done will be proportional to a^2 and hence vanish in the limit of small a. The matter action is

$$I_M = \tfrac{1}{2}\int_0^1 \left[a^2\tau^2\left(\frac{d\phi'}{d\tau}\right)^2 + 2a^4\tau^3\tilde{U}(\phi')\right] d\tau, \tag{4.12}$$

where a (which is independent of τ) and $\phi'(\tau = 1)$ are the values of a and $\tilde{\phi}$ on the bounding 3-space configuration. One can now choose $\phi'(\tau)$ to minimize I_M for fixed a and $\tilde{\phi}$. It can then be seen that the minimum $I_M(a, \tilde{\phi})$ vanishes as a goes to zero at fixed T provided that

$$\lim_{|\tilde{\phi}|\to\infty} e^{-6|\tilde{\phi}|}\tilde{U}(\tilde{\phi}) = 0. \tag{4.13}$$

We shall assume that $\tilde{U}$ obeys this condition henceforth.

If the configurations described above dominate the path integral, the wave function Ψ will tend to the same value everywhere on the light cone $T = |Z|$. We can normalize this value to be one because we are interested only in relative probabilities. We thus have the Cauchy data to solve the Wheeler-DeWitt equation (4.6) for all $a > 0$ and $\tilde{\phi}$.

For $|\tilde{\phi}|$ sufficiently large that one has

$$\tilde{U}(\tilde{\phi}) \geqslant 1, \qquad 0 < \left|\frac{d\ln\tilde{U}}{d\tilde{\phi}}\right| \ll 6 \tag{4.14}$$

for all larger values of $|\tilde{\phi}|$, the solution of (4.6) with $\Psi(0, \tilde{\phi}) = 1$ is approximately

$$\Psi = J_0(z), \tag{4.15}$$

$$z = \tfrac{1}{3}(2\tilde{U})^{1/2}a^3. \tag{4.16}$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

For large z one can write this in the WKB form

$$\Psi = \Psi_+ + \Psi_- = C\,\mathrm{e}^{iS} + \bar{C}\,\mathrm{e}^{iS}, \tag{4.17}$$

where

$$C = (2\pi z)^{-1/2}, \qquad S = \tfrac{1}{4}\pi - z. \tag{4.18}$$

The normals to the lines of constant phase S represent classical trajectories, which obey the lorentzian equations

$$\left(\frac{\sigma}{N}\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 \equiv \dot{a}^2 = a^2\left(\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}t}\right)^2 + 2a^2\tilde{U}(\tilde{\phi}) - k, \tag{4.19}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(a^3\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}t}\right) = -a^3\frac{\mathrm{d}\tilde{U}}{\mathrm{d}\tilde{\phi}}. \tag{4.20}$$

In the regime (4.14) they give a long inflationary phase during which the Universe expands roughly exponentially in time by a huge factor [6, 7, 10]. One has

$$\frac{\dot{a}}{a} \approx (2\tilde{U})^{1/2}, \tag{4.21}$$

$$\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}t} \approx -(18\tilde{U})^{-1/2}\frac{\mathrm{d}\tilde{U}}{\mathrm{d}\tilde{\phi}}. \tag{4.22}$$

One can extend the WKB solution along the classical trajectories into the region where (4.14) breaks down. The trajectories will generally oscillate in $\tilde{\phi}$ if they cross a minimum in $\tilde{U}(\tilde{\phi})$, and caustics in the WKB solution will occur where neighbouring trajectories cross near extrema of the oscillations. At each caustic a complex-conjugate pair of reflected WKB solutions is generated so that for large a and for $\tilde{\phi}$ sufficiently near a minimum of $\tilde{U}(\tilde{\phi})$, Ψ_+ will be a superposition of WKB solutions of the form (3.4).

In order to calculate the relative probability distribution of Ω from these solutions, we need to know the flux (3.8) carried along any bundle of classical trajectories. It is convenient to label the trajectories by $\tilde{\phi}_0$, the value of $\tilde{\phi}$ they have where Ψ first crosses zero as a is increased from zero. For smaller a the WKB approximation is invalid and the trajectories, which represent classical spacetimes, do not have much physical meaning. In the regime where (4.18) is valid

$$\frac{\mathrm{d}F}{\mathrm{d}\tilde{\phi}_0} = \frac{3}{2\pi}. \tag{4.23}$$

This flux per range of $\tilde{\phi}_0$ is thus independent of $\tilde{\phi}_0$ if $|\tilde{\phi}_0|$ is sufficiently large that (4.14) applies. Because the integral of (4.23) diverges as $\tilde{\phi}_0$ is taken to infinity, the probability per proper time contributed by the various classical trajectories is dominated by those coming from infinitely large $\tilde{\phi}_0$.

Once the weighting factor (4.23) has been found for each bundle of trajectories originating from large $|\tilde{\phi}_0|$, the distribution of Ω among and along the trajectories

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

must be investigated. The details of this depend upon the form of the rescaled potential $\tilde{U}(\tilde{\phi})$, but the basic behaviour simply depends upon the assumption that $\tilde{U}(\tilde{\phi})$ decreases monotonically from large $|\tilde{\phi}_0|$ to a minimum of zero at some $\tilde{\phi}$ which we may, without loss of generality, shift to be a $\tilde{\phi}=0$, and that it behaves quadratically near the minimum,

$$\tilde{U}(\tilde{\phi}) = \tfrac{1}{2}m^2\tilde{\phi}^2 + O(\tilde{\phi}^3) . \tag{4.24}$$

Then the trajectories which oscillate about this minimum behave as matter-dominated universe once the magnitude of the oscillations becomes small [6, 7, 10]. The quantity

$$E = a^3\left(\left(\frac{d\tilde{\phi}}{dt}\right)^2 + 2\tilde{U}\right) \tag{4.25}$$

is approximately conserved along such trajectories. The solution expands as a dust filled model with

$$\dot{a}^2 = Ea^{-1} - k . \tag{4.26}$$

Thus

$$\Omega = \frac{E}{E-ka} . \tag{4.27}$$

One can integrate along the normals to the surfaces of constant phase (4.16) in the inflationary regime (4.14) where E grows rapidly to estimate that

$$\ln E(\tilde{\phi}_o) \approx 18 \int_0^{\tilde{\phi}_0} \left(\frac{d\tilde{U}}{d\tilde{\phi}}\right)^{-1} \tilde{U}(\tilde{\phi})\, d\tilde{\phi} + \text{log corrections} . \tag{4.28}$$

Thus the classical universes or trajectories originating from large $|\tilde{\phi}_0|$ gain enormous "energy" E during inflation and thence spend most of their time in a matter-dominated phase with E approximately conserved.

One can compute the unconditional probability distribution of the density parameter Ω. One finds

$$P(\Omega)\, d\Omega \propto \left(\frac{\Omega-1}{k^3}\right)^{1/2} \frac{d\Omega}{\Omega^2} . \tag{4.29}$$

Ω is restricted to be greater than one in the case of $k=+1$. Eq. (4.29) then gives a normalizable probability distribution with a peak at $\Omega=\frac{4}{3}$ although the expectation value of Ω diverges. For $k=-1$, Ω is restricted to be less than one. The probability distribution (4.29) is not normalizable so the probability is entirely concentrated at $\Omega=0$. In neither case does one get a sharp prediction of $\Omega=1$.

However, it is rather meaningless to consider the unconditional relative probabilities, since they are dominated by $a>E$ at arbitrarily large E, where the energy density $\rho<E^{-2}$ is arbitarily low. No observers could exist in such regions of spacetime. The physically relevant quantity is the probability distribution for Ω

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

within some fixed interval of density $d\rho$. Solutions which start out at sufficiently large values of $|\tilde{\phi}_0|$ will all pass through a given interval $\rho \to \rho - d\rho$ in the same interval of proper time. Because the contribution to the probability is proportional to the proper time and because the probability flux is independent of $\tilde{\phi}_0$ for sufficiently large $|\tilde{\phi}_0|$, the probability distribution will be dominated by trajectories from infinitely large $|\tilde{\phi}_0|$ which will have $\Omega = 1$ when their density lies between ρ and $\rho - d\rho$. Thus the model predicts that the probability distribution of Ω is in a given density range will be entirely concentrated at $\Omega = 1$. Note that this result would not have been true had we used a different probability measure that corresponded to a hamiltonian that was conjugate to conformal time rather than proper time. The contributions of trajectories to the probability distribution of Ω would have then been proportional to the intervals of conformal time that they spent in the given density range. These intervals would have been smaller for trajectories with larger values of $|\tilde{\phi}_0|$ because they would have expanded to larger radii before they reached the given density. The probability distribution for Ω would not have been concentrated at $\Omega = 1$ but would have been spread out over a broad range.

D.N.P. expresses appreciation for the hospitality of the Department of Applied Mathematics and Theoretical Physics of the University of Cambridge, the Center for Theoretical Physics of the University of Texas at Austin and the Institute for Advanced Study. Financial support was provided in part by National Science Foundation grants PHY-8316811 and PHY-8205717 and Alfred P. Sloan.

References

[1] A.A. Starobinsky, Phys. Lett. 91B (1980) 99
[2] A.H. Guth, Phys. Rev. D23 (1981) 347
[3] The very early universe, ed. G.W. Gibbons, S.W. Hawking and S.T.C. Siklos (Cambridge University Press, 1983)
[4] S.W. Hawking, in Astrophysical cosmology, Pontificiae Academiae Scientiarum Scripta Varia, 48 (1982) p. 563
[5] J.B. Hartle and S.W. Hawking, Phys. Rev. D28 (1983) 2960
[6] S.W. Hawking, in Relativity, groups and topology II, ed. B.S. DeWitt and R. Stora (North-Holland 1984) p. 333
[7] S.W. Hawking, Nucl. Phys. B239 (1984) 257
[8] S.W. Hawking and J.C. Luttrell, Phys. Lett 143B (1984) 83
[9] S.W. Hawking and J.C. Luttrell, Nucl. Phys. B247 (1984) 250
[10] D.N. Page, *in* Quantum concepts in space and time, ed. C.J. Isham and R. Penrose (Oxford University Press, 1985) p. 274
[11] B.S. DeWitt, Phys. Rev. 160 (1967) 1113;
C.W. Misner, *in* Magic without magic: John Archibald Wheeler, ed. J.R. Klauder (Freeman, San Francisco, 1972) p. 441;
Karel Kuchař, *in* Relativity, astrophysics and cosmology, ed. W. Israel (Reidel, Dordrecht, 1973) p. 237;
M. Henneaux, M. Pilati and C. Teitelboim, Phys. Lett. 110B (1982) 123
[12] T. Christodoulakis and J. Zanelli, Phys. Rev. D29 (1984) 2738; Phys. Lett. 102A (1984) 227; ICTP preprint IC/84/48
[13] E.S. Fradkin and G.A. Vilkovisky, Phys. Rev. D8 (1973) 4241

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

FORMATION OF BLACK HOLES IN QUANTUM COSMOLOGY

L.Z. FANG and M. LI

International Centre for Theoretical Physics, I-34100 Trieste, Italy
and Centre for Astrophysics, University of Science and Technology of China, Hefei, Anhui, People's Republic of China [1]

Received 2 October 1985

The formation of black holes in the quantum cosmology scheme has been discussed by means of calculating the wave function of the universe with a black hole, which is described by a Schwarzschild-de Sitter metric. The average radius of the Schwarzschild black holes formed in the process of the birth of the universe is shown to be about $l_p^6 H^2/a^5$, where l_p is the Planck length; $\Lambda = 3H^2$ is the cosmological constant; a is the radius of the universe when it enters into the classical era.

The Hartle–Hawking quantum cosmology scheme [1] explains how the universe came from nothing. Using this method the wave functions of the universe have presently been calculated in several different cases: a conformally invariant scalar field [1]; a massive scalar field [2]; a ϕ^4 scalar field [3]. Furthermore the wave functions of the Kaluza–Klein universe have also been discussed [4].

In this letter we want to discuss the primordial formation of black holes, namely, the formation of black holes in the process of the birth of the universe. In quantum cosmology this problem is equal to finding the wave function of the universe with black holes. We consider here only the case of the universe with a black hole, which is described by a Schwarzschild–de Sitter metric. This means that we need to obtain the wave function of the universe in the minisuperspace of Schwarzschild–de Sitter metric.

The difficulty in this case is that the static form of the Schwarzschild–de Sitter metric cannot be directly used in the ground state proposal of Hartle and Hawking. We should first find an available euclidean form from Schwarzschild–de Sitter metric.

The static form of lorentzian Schwarzschild–de Sitter metric is

$$\begin{aligned} ds^2 &= H^{-2}[-(1 - r^2 - r_s/r)\,dt^2 \\ &\quad + (1 - r^2 - r_s/r)^{-1}\,dr^2 + r^2 d\Omega^2], \\ d\Omega^2 &= d\theta^2 + \sin^2\theta\, d\phi^2, \\ H^2 &= \Lambda/3, \end{aligned} \tag{1}$$

which becomes the static form of the de Sitter metric if $r_s = 0$. In comoving coordinates the de Sitter metric can be written as follows:

$$\begin{aligned} ds^2 &= H^{-2}(-dt^2 + \mathrm{ch}^2 t\, d\Theta^2), \\ d\Theta^2 &= d\chi^2 + \sin^2\chi\, d\Omega^2. \end{aligned} \tag{2}$$

By the transformation of Wick rotation $t \to i\tau$ and $\tau \to \tau + \pi/2$, the de Sitter metric can be written as

$$d\tau^2 = H^{-2}(d\tau^2 + \sin^2\tau\, d\Theta^2) \tag{3}$$

which is an euclidean geometry on S^4.

In principle, by the same method one can find an euclidean geometry from the lorentzian Schwarzschild–de Sitter metric, but it is very complicated. This problem can be simplified by solving the Einstein equation for the following euclidean metric:

$$ds^2 = H^{-2}(d\tau^2 + B^2(\tau, r)dr^2 + C^2(\tau, r)\,d\Omega^2). \tag{4}$$

In tetrad formalism the Ricci tensor of metric (4) is given by

[1] Permanent address.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

$$R_{11} = -\ddot{B}B^{-1} - 2\ddot{C}C^{-1},$$

$$R_{22} = -\ddot{B}B^{-1} - 2[(C'/B)' + \dot{B}C]B^{-1}C^{-1},$$

$$R_{33} = R_{44} = -\ddot{C}C^{-1} - [(C'/B)' + \dot{B}C]B^{-1}C^{-1} + C^{-2}[1 - \dot{C}^2 - (C'/B)^2],$$

$$R_{12} = (-\dot{C}' + \dot{B}B^{-1}C')B^{-1}C^{-1}, \quad (5)$$

in which we take $H = 1$. The Einstein equation has the form of $R_{ab} = 3\delta_{ab}$. From $R_{12} = 0$ one finds

$$C'/B = \alpha(r). \quad (6)$$

Substituting eq. (6) into $R_{11} = 3$, we have

$$\ddot{C}'/C' + 2\ddot{C}/C = -3. \quad (7)$$

Integrating eq. (7) for r, one finds

$$\ddot{C}C^2 + C^3 = r_s/2, \quad (8)$$

in which r_s is an Integral "constant". Eq. (8) has the following solution for τ:

$$\dot{C} = [a(r) - C^2 - r_s/C]^{1/2}. \quad (9)$$

$a(r)$ is also an integral "constant". From $R_{11} = R_{22}$ and $R_{33} = 3$, we have

$$-2\ddot{C}C^{-1} + C^{-2}[1 - \dot{C}^2 - (C'/B)^2] = 3,$$

or

$$\alpha^2 = -2\ddot{C}C - \dot{C}^2 - 3C^2 + 1 = 1 - a(r). \quad (10)$$

In getting the last expression we already used eqs. (8) and (9).

Without any loss of generality one can take $a(r) = r^2$, then we have a final solution

$$B = C'/(1 - r^2)^{1/2},$$

$$\int_{r_+}^{C} \frac{\mathrm{d}C}{(r^2 - C^2 - r_s/C)^{1/2}} = \tau, \quad (11)$$

in which r_+ is the smaller root of the following equation of C:

$$C^2 + r_s/C - r^2 = 0. \quad (12)$$

Obviously, in the case of $r_s = 0$, the metric (11) becomes the de Sitter one.

In semiclassical approximation the wave function is given by $\psi = A\,\mathrm{e}^{-\bar{I}}$, in which $\bar{I}$ is the action for a classical solution with the required form on the boundary of a compact four-dimensional manifold and A is a slowly varying functional with respect to classical solutions.

To obtain the wave function of the Schwarzschild–de Sitter solution, we need to evaluate the action for (11). The result is

$$-\bar{I} = \int \sqrt{g}\,(R - 2\Lambda)\,\mathrm{d}^4X + \text{surface term}$$

$$= \frac{8\pi}{H^2} \int (1 - r^2)^{-1/2}[2rC - 6C^2\dot{C} + 2\dot{C}]\,\mathrm{d}r\,\mathrm{d}\tau, \quad (13)$$

in which the integration of r covers the range from $r_{\min} = \sqrt{3}(r_s/2)^{1/2}$ to 1. $r_{\min}$ is determined by eq. (12).

In the first-order approximation of r_s, one can find from eq. (11) that

$$C = r\sin\tau + (r_s/r^2)\,f(\tau),$$

$$f(\tau) = -\tfrac{1}{2} + \tfrac{3}{2}\cos\tau + \tfrac{1}{2}\cos\tau\ln(r_s/2r^3) + \tfrac{1}{4}\cos\tau\ln[(1 + \cos\tau)/(1 - \cos\tau)]. \quad (14)$$

Therefore, in this approximation, the action (13) can be expressed as

$$-\bar{I} = -\bar{I}_0 + r_s S_1(\tau) + r_s \ln r_s S_2(\tau) + r_s(\ln r_s)^2 S_3(\tau) + (\tfrac{1}{2}r_s)^{1/3} S_4(\tau) + (\tfrac{1}{2}r_s)^{1/3}\ln r_s S_5(\tau),$$

$$-\bar{I}_0 = 8\pi^2(1 - \cos^3\tau), \quad (15)$$

where

$$S_1 = (8/\sqrt{3})\pi - (8/\sqrt{3})\pi\cos\tau(2 - \cos^2\tau) - 4\pi D\sin\tau + 8\pi EF + \tfrac{16}{3}\pi\ln\tfrac{27}{256}G,$$

$$S_2 = 16\pi(\ln 27/2^{1/3}8 - 2)\sin\tau + 16\pi\ln\tfrac{27}{256}F + \tfrac{32}{3}\pi G,$$

$$S_3 = -\tfrac{32}{3}\pi\sin\tau + \tfrac{64}{3}\pi F, \quad (16)$$

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

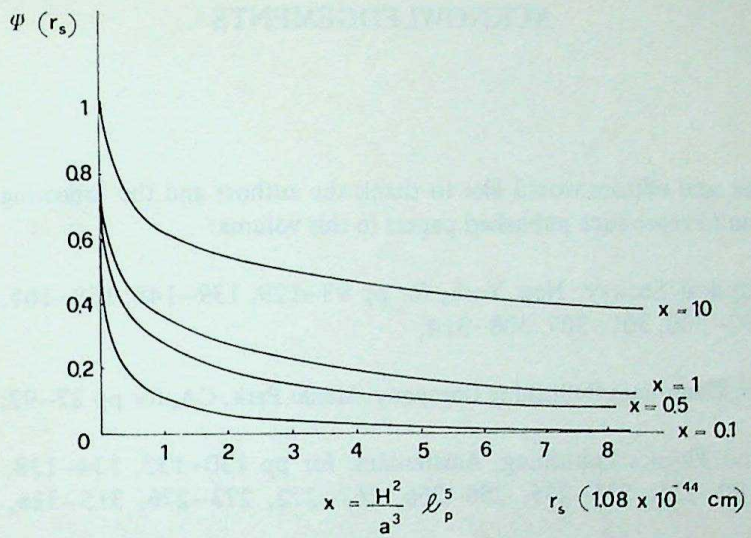


Fig. 1. The wave function of the Schwarzschild–de Sitter universe.

$$S_4 = \tfrac{64}{3}\pi(\ln\tfrac{9}{2} - 4)\sin\tau + 16\sqrt{3}\,\pi(\cos\tau - 1) - \tfrac{32}{3}\pi\sin\tau\ln[(1+\cos\tau)/(1-\cos\tau)],$$

$$S_5 = \tfrac{64}{3}\pi\sin\tau, \qquad (16\text{ cont'd})$$

and

$$D = \tfrac{5}{2}\pi^2 + \ln\tfrac{27}{256}(\ln 432 - 4) + 12,$$

$$E = \tfrac{1}{2}\pi^2 + \tfrac{1}{3}\ln\tfrac{27}{256}\ln 432,$$

$$F = \sin^2\tau\,(\tfrac{3}{2} - \sin\tau),$$

$$G = \tfrac{1}{2}\sin\tau\{1 - \ln 2 - \cos\tau + \tau/\sin\tau + \tfrac{1}{2}\ln[(1+\cos\tau)/(1-\cos\tau)]\}. \qquad (17)$$

Eqs. (15)–(17) are valid in the range $a = H^{-1}\sin\tau \leqslant H^{-1}$ or $Ha \leqslant 1$. The classical era of the universe corresponds to the range $Ha \geqslant 1$, in which a is the radius of the universe. The wave function in the range $Ha > 1$ can be found by the joint condition of the WKB method.

The formation of a black hole is described by the behaviour of the wave function with respect to r_s, $\psi(r_s)$, which is shown in fig. 1. We can see from fig. 1 that the amplitude of the wave function decreases quickly with the increase of r_s. This may show that a black hole of considerable size cannot form in the beginning of the universe.

The dominant term in $\psi(r_s)$ is

$$\exp[-(r_s/2)^{1/3}\ln(1/r_s)S_5], \qquad (18)$$

where $S_5 = \frac{64}{3}\pi\,Ha$. Therefore, the average of r_s is about S_5^{-3}. Considering $H^2 l_p^2 = 1$ and the dimensional Schwarzschild radius of the black hole to be equal to $r_s H^{-1}$, one finally finds the average of r_s:

$$\bar{r}_s \sim H^2 l_p^6/a^3. \qquad (19)$$

Eq. (19) is valid in both cases, $Ha < 1$ and $Ha > 1$.

The authors would like to thank Professor Abdus Salam, the International Atomic Energy Agency and UNESCO for hospitality at the International Center for Theoretical Physics, Trieste.

References

[1] J.B. Hartle and S.W. Hawking, Phys. Rev. D28 (1983) 2960.
[2] S.W. Hawking and Z.C. Wu, Phys. Lett. 151B (1985) 15.
[3] I.G. Moss and W.A. Wright, Phys. Rev. D29 (1984) 1067.
[4] Z.C. Wu, Phys. Lett. 146B (1984) 307; Z.C. Wu and X.M. Hu, Phys. Lett. 149B (1984) 87.

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

ACKNOWLEDGEMENTS

The Publisher and editors would like to thank the authors and the following publishers for permission to reproduce published papers in this volume:

American Physical Society, New York, for pp 93–129, 139–148, 158–161, 174–189, 277–291, 292–300, 301–307, 308–314;

The Benjamin/Cummings Publishing Company, Menlo Park, CA, for pp 27–92;

North-Holland Physics Publishing, Amsterdam, for pp 130–133, 134–138, 149–153, 154–157, 190–235, 236–255, 256–266, 267–272, 273–276, 315–326, 327–329;

Specola Vaticana, Citta del Vatticano, for pp 162–173.

...College of Education
Raipur, Bantalab
Jammu.
Acc No....5263....
Dated....19/3/2024

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

Gandhi Memorial College Of Education Bantalab Jammu

CC-0 Agamnigam Digital Preservation Foundation,Chandigarh.Funding by IKS-MoE-2025-Grant

0407 sc

ISBN 9971-50-312-3 pbk